

# nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

## BEYOND MOORE'S LAW

Why its approaching end  
is a new beginning for the  
computer industry

PAGE 144

GETTING PUBLISHED

### WORTH THE WAIT?

The pain of review,  
re-review and re-re-review

PAGE 148

CLIMATE CHANGE

### NO EASY ANSWER

CO<sub>2</sub> removal may cause as  
many problems as it solves

PAGE 153

HOMINID EVOLUTION

### A PARTING OF THE WAYS

The gorilla-human  
lineage split revisited

PAGE 215

NATURE.COM/NATURE

11 February 2016 £10

Vol 530, No. 7589



# THIS WEEK

## EDITORIALS

**CARBON** End the one-hit wonders of climate-change action **p.130**

**WORLD VIEW** Structural changes to tackle sexual harassment **p.131**



**TRUMPET NOSE** Fossil shows animal had handy hidden hooter **p.133**

## Benefits of sharing

*A swift and effective response to emerging infectious diseases demands that researchers have ready access to the latest data on the pathogens responsible. There is still a long way to go to ensure this.*

Another year, another virus. As the Ebola-virus epidemic recedes, Zika dominates the news. The virus, which usually causes only mild symptoms, has been linked to a reported increase in the number of babies born in Brazil with microcephaly — abnormally small heads and brains. The possible implications of this for pregnant women demand a rapid and evidence-based approach.

The immediate priorities are to gather epidemiological and clinical data to establish whether the apparent spike in cases is real and, if so, to what extent the Zika virus is involved (see page 142). And researchers elsewhere must have full access to all of this information as soon as it is available.

Conventional scientific publishing, based on rounds of peer review, can be too slow to rapidly disseminate research findings during a public-health emergency. One solution is the immediate release of data to public databases and subsequent publication of peer-reviewed analysis. As we have said before, prior release of data and analysis to public databases, preprint servers and forums will not jeopardize consideration of a submission to Nature journals. And Nature journals will make all papers relating to Zika virus free to access until further notice.

Already, there have been promising moves to make data on microcephaly, and on the epidemic of Zika in the Americas, readily accessible. The World Health Organization (WHO) has announced a 'Zika Open' initiative, in which all relevant submissions to its *Bulletin* will be posted online within 24 hours.

In this issue, we publish research that demonstrates the need for rapid data sharing during outbreaks. As detailed on page 228, genome-sequencing technology has advanced to the point at which the whole genome of a virus sample can be sequenced in the field within 24 hours by means of a portable sequencing system. Previously, such sequencing during the course of an epidemic has been slower because it has relied on sending samples to a laboratory. Although this method still faces technical challenges, it should prove a crucial tool in epidemiological research. It offers the potential to rapidly trace how a disease is being passed from person to person, and so help to guide authorities to direct resources that can break these transmission chains.

But the full potential of this advance can be realized only if scientists can access sequence data obtained from samples taken at different times and places during an outbreak. Pathogen genome sequences are most useful when studied alongside epidemiological and clinical data.

Many scientists across the research fields, but notably in genomics, have been enthusiastic champions of early data release. Infectious-disease researchers make use of public forums and databases such as virological.org and GISAID (Global Initiative on Sharing All Influenza Data). The former resource is becoming a major platform for sharing and discussing preliminary analyses of data and it already hosts a genomic analysis of the Zika virus. And, launched just over four years ago, ISARIC (International Severe Acute Respiratory and Emerging Infection Consortium) works with clinicians and epidemiologists to

put in place pre-agreed protocols and data-sharing processes that can then quickly be adapted to a new situation.

Almost one year ago, this journal published a Comment article that called for the immediate sharing of outbreak data, a policy that the authors themselves had adopted during their early sequencing of Ebola-virus genomes in the 2014–15 outbreak (N. L. Yozwiak *et al.* *Nature* **518**, 477–479; 2015). They also called on the WHO to convene a meeting to develop guidance for data sharing during infectious-disease outbreaks. Such a meeting took place last September. It was attended by government representatives, public-health agencies, scientists, research funders, ethicists and publishers. All acknowledged that pathogen sequence information collected during a public-health emergency is of greatest value when released openly, in as close to real time as possible.

A statement released by the WHO after the meeting emphasized the "fundamental moral obligation" of every researcher who generates information related to a public-health emergency to share their preliminary results once these have undergone quality control. Representatives from leading biomedical journals also unequivocally emphasized that disclosure of such information would not prejudice journal publication.

Scientists still face challenges to swift data sharing. For example, as was seen during the emergence of the H5N1 and H7N9 avian influenza viruses and Middle East respiratory syndrome coronavirus, rapid data sharing can be hampered by a lack of international rules that govern how credit and rights (including intellectual property) should be fairly distributed among scientists and authorities in the countries where outbreaks occur as well as researchers based elsewhere.

To play our part in driving the shift towards fast data sharing during public-health emergencies, Nature journals will now encourage authors who haven't already deposited their relevant sequence information in public archives to do so on submission. ■

## A good precedent

*Jimmy Carter's efforts to eradicate Guinea worm should be applauded.*

When former US President Jimmy Carter wanted Ghana to take his goal to eradicate the Guinea worm seriously, he came up with a novel threat. Carter told the country's president that he would try to get the parasitic disease's name changed to Ghana worm. "There isn't a Guinea worm left in Ghana now," Carter told journalists with a grin in London last week.

Carter might just be the United States' most productive ex-president. In the 35 years since he left the White House, the peanut farmer from



Georgia has brokered peace deals, helped to see off despots such as Panama's Manuel Noriega and still found time to spend one week each year building houses for the needy.

But Carter's most lasting accomplishment is likely to be disease eradication. Last year, just 22 people in 4 countries in sub-Saharan Africa had Guinea-worm disease, compared with the 3.5 million who suffered the parasitic infection in one year in the mid-1980s, when the non-profit Carter Center in Atlanta, Georgia, began to lead a global campaign to wipe out the disease.

The former president spoke last week at a meeting to drum up support for Guinea-worm-eradication efforts, barely showing his 91 years or his fight with cancer (he revealed last year that he has metastatic melanoma). He has said that he wants the last Guinea worm to leave the world before he does, and he may also live long enough to see the back of another global menace: the three-decade effort to eradicate polio is nearing completion. The virus is now present in only Afghanistan and Pakistan, which together recorded an all-time low of 73 cases last year.

Neither disease will go without a final struggle. Violence and political instability threaten both efforts, and a Guinea-worm epidemic in dogs in Chad demonstrates the unpredictability of disease eradication. But both campaigns have come too far and cost too much to be left incomplete, and world leaders should provide the resources needed to finish the jobs.

What next? Carter is among those who would like to see other diseases exterminated. His centre launched the International Task Force for Disease Eradication, which has drawn up a list of candidates, including measles, mumps and lymphatic filariasis (a parasitic infection also known as elephantiasis). "We're ready and eager to go on several diseases," Carter said. He even needed the World Health Organization for its sluggishness in adopting Guinea-worm eradication as an official goal, and its reluctance to target other diseases for eradication.

Carter might be right on Guinea worm, but given the tools available to twenty-first-century science and medicine, is an all-out assault to eradicate disease still the best way to go?

Putting the fight against a pathogen on a pedestal can warp efforts to tackle other diseases and develop health systems. Amid regular

door-to-door campaigns to deliver oral polio vaccines, northern Nigeria had (and still has) low rates of routine childhood vaccination for diseases such as measles and diphtheria. (Officials have argued that future eradication campaigns could do more to improve health systems, for instance by starting eradication efforts in a particular country only once routine immunization coverage reaches a certain threshold.)

The Guinea-worm and polio campaigns have been decidedly old-school in approach. The main tool against polio has been a decades-old oral vaccine (although a switch to an injectable vaccine is under way). There is no vaccine or treatment for Guinea worm, and eradication efforts have focused on community-based approaches to change behaviour, such as teaching people to filter all drinking water, to avoid recontaminating water sources, and to report cases to health authorities.

Future eradication campaigns have obvious appeal. Although they are costly, success spares countries from the economic burdens of a disease. Guinea-worm efforts have also helped to provide clean drinking water to millions of people in some of the world's poorest and most remote areas. And eradication is an appealing target that can court donors who might not otherwise support public health.

Future eradication campaigns (and other public-health efforts) should find inspiration in the fact that Guinea-worm eradication has been focused on behavioural change, but they are likely to have other approaches to choose from. One of the most promising is gene-drive technology, which allows a gene that is harmful to a pathogen or vector to rapidly spread through a sexually reproducing population.

Last year, two teams developed experimental gene drives that made mosquito populations infertile or resistant to malaria. It will be years before the technology is ready for field trials. But it is one reason that people are beginning to talk seriously again about malaria eradication, after failed efforts in the 1950s and 1960s.

Gene drives could also be applied to other vector-borne infections, including tick-borne infections such as the bacteria that cause Lyme disease. President Carter's hit list may yet grow longer. ■

**"Eradication success spares countries from the economic burdens of a disease."**

## Outside the bubble

*Governments must stop proposing solutions and invest in large-scale removal of carbon dioxide.*

Every generation throws a different hero up the pop charts, sang musician Paul Simon. And every political cycle, it seems, promotes its own signature solution to the problem of how to curb and prevent climate change. This year's answer is bioenergy with carbon capture and storage — a plan on a colossal scale to grow grass and trees on an area half the size of the United States, harvest and ship this biomass to power stations, burn it and then trap the carbon dioxide from the exhaust gases. The greenhouse gas would then be piped underground and stored indefinitely. The scheme's acronym — BECCS — barely does its complexity justice.

Perhaps this is the first you have heard of BECCS. That wouldn't be a surprise. It is something of an overnight sensation — the boy band launched to number one on the back of a reality television show, rather than the grizzled rockers who earned their fame after years of concerts attended by three people and a dog. Yet at the Paris climate talks late last year, which were widely acclaimed as a triumph, the BECCS scheme was quietly installed as the world's Plan A.

That's because it comes with a very catchy tune that politicians can't get out of their heads. BECCS solves the problem of future carbon emissions and cleans up the past. The plants suck CO<sub>2</sub> from the atmosphere,

which ends up safely underground. We get the benefit of burning them to generate electricity, the world gets to keep its power infrastructure, and the atmosphere experiences what BECCS enthusiasts call negative emissions. Rather brilliantly, the more energy the scheme produces, the more the planet edges away from dangerous levels of global warming.

What's not to like? Well, in a World View article published after the Paris talks, climate scientist Kevin Anderson compared BECCS to a fairy godmother, conjured up to wish away reality in a puff of optimism (see *Nature* 528, 437; 2015). And in a Comment piece on page 153, environmental scientist Phil Williamson takes a hard look at some of the questions that BECCS seems to pose, and finds few answers.

How would we preserve forests and grasslands, faced with such a demand for energy crops? How much carbon would be released during the agricultural stage? How much water will we need, and where will we get it? How much will it cost to build the network of compressors, pipes, pumps and tanks that will be needed to liquefy and transport the separated CO<sub>2</sub>? Can it even be separated at a sensible cost?

Recent years have seen a series of solutions to global warming thrown up the political agenda. From biofuels and carbon offsets to ocean fertilization and conventional carbon capture, each has had its moment in the sun, only to be replaced by something younger with a new sound.

BECCS may yet prove to have staying power. But to avoid another one-hit climate wonder, governments must spend the money to do the groundwork — as Williamson says — and to answer those questions and plenty more. Politics is the art of the possible. But serious action on climate change must be based on the science of the probable. ■

➔ **NATURE.COM**  
To comment online,  
click on Editorials at:  
[go.nature.com/xhunqv](http://go.nature.com/xhunqv)



## Change the system to halt harassment

*Universities and their senior staff must do more to deter, detect and punish all forms of inappropriate behaviour, says Joan Schmelz.*

Sexual harassment is a vile plague on science. I know this because I am asked regularly for help and advice by women who have been harassed. And the problem has been laid out for the broader research community in a series of recent cases. The US National Science Foundation, among other organizations, has reminded universities of their obligations to investigate allegations of harassment and to punish perpetrators accordingly. Yet the system continues to protect the reputation of harassers and the institutions that employ them. It is not enough for affected individuals to speak up. They must also know that their complaint will be dealt with properly. Here are some suggestions that could improve this situation.

Everyone should know the law. Where sexual harassment is illegal, the bar for inappropriate behaviour is lower than some might think. In the United States, for example, harassment constitutes behaviour that is “sufficiently severe, persistent, or pervasive to limit a student’s ability to participate in or benefit from an educational program or activity, or to create a hostile or abusive educational environment,” according to the Department of Education.

Students, members of faculty and other university staff sometimes misinterpret these words to mean that severe, persistent or pervasive behaviours define a hostile environment. They do not understand that these terms can stand alone. Behaviour that many people might regard as being innocuous — but that is persistent, targeted or timed to have a negative effect — can limit the ability of a student to learn and is unacceptable.

Every university should have an office to which members of its community can talk anonymously about harassment — an ‘Office of Good Advice’. It must be fundamentally separate from the affirmative-action office, legal team and police force of the university — all of which are obliged to report apparent breaches of the law.

The Office of Good Advice should be well known to all. It should be staffed by professionals, and it should be the first result that is returned in a web search for ‘sexual harassment’. Anyone on campus who needs to talk about harassment should know which university employees will report incidents and which can keep such reports confidential.

Every university needs an Office of Good Advice. This is so that students need not fear that they will be pressed to make a formal, legally viable report at offices with legal responsibility, and because staff who are released from legal obligations are more likely to be able to listen with objectivity and sensitivity. Those who have been harassed might later report that they were asked intimidating and inappropriate questions that seem to undermine the validity of their complaint, such as: “Were you drinking?” or “Are you unhappy with your grade in his course?” No matter how good their

intentions, members of staff in offices who must uphold the law cannot help but be influenced by this responsibility and their knowledge of the requirements of an investigation.

The Office of Good Advice needs to keep track of multiple complaints against the same harasser. A university, by its very nature, has a fluid population of students and postdoctoral researchers. A way in which serial harassers can continue is by targeting one individual at a time, perhaps in a different class or a year apart, so that their targets never meet and are unaware of each other.

At present, reports of sexual harassment are filed under the name of the complainant, in part to protect the identity of the alleged harasser. One small step in the right direction would be to file the report under the name of the harasser. There is reason to be cautious, but the law — at least in the United States — makes it clear that universities should have

the flexibility to investigate allegations thoroughly.

Department heads must step up. They have the power to change the culture and to create an environment in which everyone feels able to do their best work. In fact, it is a fundamental component of their job description. Although they should embrace their responsibility, too many senior staff choose not to. They convince themselves that they are too busy: after all, they have reports to file, budgets to balance, proposals to write, postdocs to hire and students to supervise.

These senior figures should make themselves and their staff familiar with the local anti-harassment policy and available resources. They should post the policy and information on how to access the resources in a prominent place. Such actions

clearly demonstrate that the department considers the issue to be important and takes it seriously.

Department-wide meetings on the topic of sexual harassment should also be convened, with help from on-campus resources. The chair of the meeting and the university hierarchy should affirm that harassment will not be tolerated. The existing policy and resources should be discussed. Everyone should be encouraged to share their views on the workplace culture — with the group, in smaller, targeted meetings or anonymously.

What can everyone else do? We can project that we are safe to come to for advice, simply by interjecting with “I disagree”, when inappropriate or sexist comments arise in group discussion. This small phrase can go a long way. And if someone comes to you with a problem, do not immediately question their story. Instead, listen respectfully. You can be certain that if action is going to be taken, the appropriate questions will be asked in due course. ■

**Joan Schmelz** works for the Universities Space Research Association and is deputy director of the Arecibo Observatory in Puerto Rico. e-mail: jschmelz@usra.edu

EVERYONE  
SHOULD BE  
**ENCOURAGED**  
TO SHARE THEIR  
VIEWS ON THE  
**WORKPLACE**  
CULTURE.

➔ **NATURE.COM**  
Discuss this article  
online at:  
[go.nature.com/7leinz](http://go.nature.com/7leinz)



# RESEARCH HIGHLIGHTS

Selections from the  
scientific literature

## GENOMICS

### Early risers share genetic signature

Three different teams have identified genetic variants that are associated with being a morning person.

Youna Hu, David Hinds and their team at 23andMe, a personal-genetics firm in Mountain View, California, analysed genome data from nearly 90,000 customers who stated in a survey whether they were a morning person or a night owl. Teams led by Michael Weedon at the University of Exeter Medical School, UK, and Richa Saxena at the Massachusetts General Hospital in Boston conducted similar analyses on the DNA of more than 100,000 people who participated in the UK Biobank project.

Each team found a dozen or more variants that were linked to a person's preference for sleep times — and many were in or near genes that help to regulate the body's circadian clock.

*Nature Commun.* 7, 10448 (2016); preprints at bioRxiv <http://doi.org/bcb7> (2016); <http://doi.org/bcb8> (2016)

## CLIMATE CHANGE

### More carbon from planted forests

Planting trees will not necessarily slow climate change.

Kim Naudts at the Laboratory of Climate Science and Environment in Gif-sur-Yvette, France, and her colleagues paired a history of land-use in Europe with land and atmospheric models to study the effect of forests on the climate. Although the continent's forests have expanded by 10%, timber harvesting and a shift to more

commercially valuable trees — mainly the fast-growing conifers — have resulted in the release of more than 3 billion tonnes of carbon into the atmosphere since 1750.

The change from deciduous trees to darker-leaved conifers contributed to a rise of 0.12 °C in local surface temperatures. *Science* 351, 597–600 (2016)

## NEUROSCIENCE

### Brain cells in wells make amyloid

Many different types of brain cell can secrete high levels of the peptide amyloid- $\beta$ , which forms the brain plaques associated with

Alzheimer's disease.

To detect molecules from single brain cells, Tracy Young-Pearse at Harvard Medical School in Boston, Massachusetts, and her colleagues designed arrays of nanometre-sized wells. They took cells from healthy people and from individuals with familial Alzheimer's disease, reprogrammed them into stem cells and used them to derive a variety of brain-cell types, which they then placed in the wells. The authors added antibodies that would detect specific molecules.

Subsets of the brain cells secreted amyloid- $\beta$  at different rates, and even non-neuronal cells such as

evolutionary and geographical history. The team found that the virus has spread from Europe in the past century, in part because of the trade in commercial colonies. New carriers of the virus also contributed: the *Varroa* mite (*Varroa destructor*) began to infect European honeybees about 50 years ago, when this pandemic started.

Tighter controls on commercial colonies are needed to slow the virus's spread, the authors say. *Science* 351, 594–597 (2016)



## PARASITOLOGY

### Honeybee virus spread by human activity

The commercial trade in honeybees has helped to spread a deadly bee virus around the globe over the past century.

Deformed wing virus reduces the winter survival of European honeybees (*Apis mellifera*), and could be a factor in the large colony losses seen in some parts of the world. To find out how the virus became pandemic, Lena Wilfert at the University of Exeter, UK, and her colleagues analysed the virus's genome to reconstruct its

astrocytes produced high levels. More of the cells derived from people with Alzheimer's disease secreted large amounts of the peptide compared to cells from healthy individuals.

*J. Neurosci.* 36, 1730–1746 (2016)

## NEUROSCIENCE

### Molecule protects ageing neurons

A small molecule reverses the neuronal deterioration that is typically found in middle-aged animals.

Compared with the neurons of young animals, nerve cells in middle-aged rats generally have fewer branches. Their

STEPHEN MARTIN, UNIV. SALFORD

connections, or synapses, also seem to be less able to strengthen with repeated activation — changes that could help to explain declines in learning and memory over time. Christine Gall, Gary Lynch and their colleagues at the University of California, Irvine, preserved neuronal branching and boosted synapse strengthening in rats by treating them with a molecule that enhanced the activity of the AMPA-type glutamate receptor. This increased excitatory brain signalling. Animals treated for three months also showed improvements in some tests of long-term memory.

The authors suggest that the mechanism could one day be used to protect against age-related neuronal decline. *J. Neurosci.* 36, 1636–1646 (2016)

## PALAEOLOGY

## Mammal with a dinosaur nose

A wildebeest-like creature that lived in Africa 70,000 years ago had a trumpet-shaped nasal passage, similar in shape to those of some dinosaurs.

*Rusingoryx atopocranium* was previously known from incomplete specimens, but a team led by Haley O'Brien at Ohio University in Athens and Tyler Faith at the University of Queensland in Brisbane, Australia, have discovered six more-complete skulls on an island in Lake Victoria in Kenya (one of the skulls is pictured). X-ray scans revealed large, hollow nasal passages that extended out from the top of the animals' heads. Known as nasal domes,



these have previously been seen only in some hadrosaurs that lived around 65 million years ago — a rare example of how similarities in behaviour and environment can spur the evolution of a common feature in distantly related animals.

Acoustic modelling suggested that the structures allowed *R. atopocranium* to vocalize at very low frequencies, possibly helping herds to communicate without alerting predators, the authors propose.

*Curr. Biol.* <http://doi.org/bcdn> (2016)

## LINGUISTICS

## Languages have common structure

Many languages share a universal semantic structure that is independent of their speakers' culture, environment or how closely the languages are related.

Hyejin Youn and Tanmoy Bhattacharya at the Santa Fe Institute in New Mexico and their colleagues studied the words for 22 universal concepts — such as 'Sun' and 'water' — across 81 languages, and identified which words had multiple meanings. They linked related words and meanings together to form network maps, and found common patterns in the way that different meanings were connected to each other. For example, words for 'sea' and 'salt' were closer to each other than to words for 'Sun', and this structure was preserved across all the languages.

This method could help to reveal concepts that are universal features of human cognition and language use, the authors say.

*Proc. Natl Acad. Sci. USA* <http://doi.org/bb94> (2016)

## MATERIALS

## Better battery with more juice

Researchers have made a battery that can both hold high energy and deliver high power.

## SOCIAL SELECTION

Popular topics on social media

### 'A field of crop scientists'

When cancer researcher Ritankar Majumdar published a blog post about the collective names of doctors and scientists on 2 February ([go.nature.com/fzcnqt](http://go.nature.com/fzcnqt)), he had no idea he would inspire #scientistherdnames. Hundreds of tweets with the hashtag, including "a cloud of data scientists" and "a nucleus of physicists", soon dominated scientists' Twitter streams. Some tweets reflected the angst experienced in the lab. Erol Akçay, a theoretical biologist at the University of Pennsylvania in Philadelphia, tweeted: "A grant of silent sobs."

➔ **NATURE.COM**  
For more on popular papers:  
[go.nature.com/8njfmx](http://go.nature.com/8njfmx)

Batteries typically can hold a lot of energy but discharge it at low rates. To make a material capable of both high energy and high power, a team led by Yan Yu at the University of Science and Technology of China in Hefei placed sodium vanadium triphosphate ( $\text{Na}_3\text{V}_2(\text{PO}_4)_3$ ) inside a mixture of reduced graphene oxide and carbon nanotubes to improve the sodium compound's conductivity. They deposited the material onto a current-collecting stainless steel support, avoiding the use of additives that would reduce the performance of the device.

The team made a battery with the porous material and showed that the device could deliver high levels of energy at a high rate. It also kept 96% of its capacity even after being charged and discharged 2,000 times. The approach could be used with other electrode materials, the authors say.

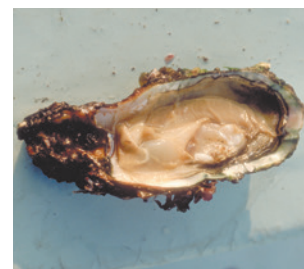
*Adv. Mater.* <http://doi.org/f3kvx3> (2016)

## ENVIRONMENTAL SCIENCE

## Ocean plastic hurts oysters

Oysters that consume the small pieces of plastic littering the world's oceans produce fewer and less-robust offspring than their unexposed counterparts.

Millions of tonnes of plastic end up in the oceans every year, and tiny 'microplastic'



fragments — those smaller than 5 millimetres in diameter — could be harming marine life. Arnaud Huvet at the French marine-research agency IFREMER in Plouzané and his colleagues placed Pacific oysters (*Crassostrea gigas*; pictured) in water laced with micrometre-sized polystyrene spheres, at levels estimated to be similar to those where oysters live in the wild. After two months, oysters exposed to the plastic produced fewer and smaller egg cells, less-mobile sperm and fewer offspring than did animals raised in water without the plastic. The offspring also grew more slowly.

Ingested plastic might be disturbing the oysters' digestion and releasing endocrine-disrupting chemicals, which affect reproductive systems, the authors say.

*Proc. Natl Acad. Sci. USA* <http://doi.org/bcdm> (2016)

➔ **NATURE.COM**  
For the latest research published by Nature visit:  
[www.nature.com/latestresearch](http://www.nature.com/latestresearch)



# SEVEN DAYS

The news in brief

## FUNDING

### Zika funding boost

US President Barack Obama requested US\$1.8 billion in emergency funding to fight Zika virus, the White House announced on 8 February. The mosquito-transmitted virus might be linked to head malformations called microcephaly in newborn infants, and has mostly affected countries in South America. The World Health Organization has declared the rise in microcephaly a global health emergency. The president's proposal is part of his 9 February US budget request, and would give around \$1.5 billion to the Department of Health and Human Services, including \$828 million for the US Centers for Disease Control and Prevention to research the virus, and about \$200 million to develop a vaccine. The US Agency for International Development will get \$335 million.

## EVENTS

### Copenhagen cuts

Government budget cuts in Denmark have forced the University of Copenhagen to slash staff numbers. More than 500 employees will lose their jobs, one-quarter of them researchers, according to a 4 February news release from the university. Voluntary redundancies and discontinued contracts will make up 323 of the job losses, and 209 staff can expect to be laid off. The university will also reduce its PhD intake by 10%. The job losses come after the Danish government cut funding for research by 1.4 billion kroner (US\$210 million) in October.

### Climate-job losses

Hundreds of climate researchers at Australia's national science agency, the



## Reactor roars into life

German Chancellor Angela Merkel ignited the first hydrogen plasma at the Wendelstein 7-X stellarator in Greifswald on 3 February, marking the device's official launch. The fusion reactor's design differs from that of the other most-promising fusion facility, ITER, being built near Cadarache, France. The official switch-on produced a hydrogen plasma for just a fraction of a second, but the helical-shaped stellarator should be able to operate continuously — a major advantage over the pulsed generation of ITER's tokamak design.

Commonwealth Scientific and Industrial Research Organisation (CSIRO), are set to lose their jobs. Chief executive Larry Marshall said on 4 February that the agency would be shedding up to 350 jobs over the next two years as it shifts its strategy away from basic climate science. CSIRO, which employs more than 5,000 people, had already dropped 1,300 jobs in the 2 years up to June 2015, largely because the 2014 federal budget cut the agency's funding by Aus\$115 million

(US\$108 million) over 4 years. See [go.nature.com/apbkmk](http://go.nature.com/apbkmk) for more.

### Drug-trial probe

A preliminary investigation into the death of a participant in a first-in-human drug trial in France, and the hospitalization of five others, has failed to establish a cause. In a report to health minister Marisol Touraine on 4 February, the general inspectorate for social affairs criticized trial contractor Biotrial for giving doses of the test drug to other

subjects on 11 January, despite the first subject having been hospitalized the previous day. The trial was halted later that day, and the report says that Biotrial respected trial protocols. The five patients have since been discharged from hospital.

### Methane lawsuit

The state of California announced on 2 February that it had filed a lawsuit against the Southern California Gas Company for failing to promptly report and halt emissions from a massive natural-gas leak at one of its storage facilities. Discovered on 23 October last year, the gas leak at the Aliso Canyon natural-gas storage facility has yet to be controlled. As of 8 January, cumulative methane emissions from the facility were estimated to be equivalent to more than 2 million tonnes of carbon dioxide, or 2% of the state's greenhouse-gas emissions over the 11-week period.

## BUSINESS

### CRISPR goes public

Editas Medicine, a company in Cambridge, Massachusetts, that aims to create therapies from CRISPR-Cas9 gene-editing technology, went public on 2 February, raising US\$94.4 million. The initial public offering was seen as a vote of confidence both in the sluggish biotechnology market and in a tool that has yet to reach clinical testing. Several pioneers in CRISPR-Cas9 technology are among the founders of the company.

### China chemicals bid

China National Chemical (ChemChina) has offered to buy agrochemicals firm Syngenta in a US\$43-billion deal that could alter the

landscape of the industry. Syngenta, headquartered in Basel, Switzerland, is best known for its agricultural seed and pesticides business, and was wooed by its competitor Monsanto in a takeover bid last year. But on 3 February, Syngenta announced that it had accepted an offer from ChemChina of Beijing. The deal must be approved by regulators before it can be completed.

## PEOPLE

## HHMI president

The Howard Hughes Medical Institute (HHMI) in Chevy Chase, Maryland, has appointed its first female president. Erin O'Shea, a systems biologist from Harvard University in Cambridge, Massachusetts, will take office in September, the organization announced on 3 February. O'Shea, whose speciality is gene regulatory networks, has served as HHMI's chief scientific officer and vice-president since 2013. She will be HHMI's sixth president, replacing biochemist Robert Tijan of the University of California, Berkeley.

## Macchiarini out

On 4 February, the Karolinska Institute (KI) in Stockholm announced that it will not renew the



contract of surgeon Paolo Macchiarini (pictured) when it ends in November 2016. The KI visiting professor pioneered artificial-windpipe transplants. A Swedish documentary alleged that he broke ethical guidelines, prompting the KI to consider reopening an investigation into Macchiarini's work and to ask him to phase out his research. The KI has also ordered an external investigation into the way the case was handled. Urban Lendahl, a geneticist at the KI, resigned as secretary-general of the Nobel Assembly on 6 February because he expects to become involved in the investigations. See [go.nature.com/rh1ne4](http://go.nature.com/rh1ne4) for more.

## RESEARCH

## EMBO in India

Europe's life-sciences organization EMBO has extended its reach in Asia. On 4 February, India became

EMBO's second Associate Member State. Researchers in India are now eligible to participate in all EMBO programmes, including courses and workshops, long- and short-term fellowships and the EMBO Young Investigator Programme. Singapore became the organization's first Associate Member State last July.

## Italian misconduct

The ethics committee of Italy's National Research Council, the CNR, announced on 9 February that it has identified 17 cases of scientific misconduct in published papers co-authored by researchers in CNR labs. The CNR did not publicly identify the papers, but says that it is discussing next actions with the relevant authors. The committee assumed responsibility for research integrity two years ago, and is now examining a further 61 cases. It published guidelines for good scientific practice in June last year, committing it to investigating all allegations of misconduct brought to its attention.

## Smoking laws

Smoking bans around the world have measurably improved the health of non-smokers, according to a 4 February review (K. Frazer *et al.* *Cochrane Database Syst.*

## COMING UP

## 11 FEBRUARY

A hotly anticipated announcement from the LIGO collaboration about the existence of gravitational waves, which were predicted 100 years ago by Albert Einstein, will be made at the National Press Club in Washington DC. [go.nature.com/cqt5te](http://go.nature.com/cqt5te)

## 12 FEBRUARY

Astro-H, an X-ray astronomy satellite led by the Japanese Space Agency, will launch from Tanegashima Space Center, Japan (see page 140 for more).

## 14–19 FEBRUARY

Whitefly, which can devastate crops, is discussed at the 2nd International Whitefly Symposium in Arusha, Tanzania. [www.iita.org/iws2016](http://www.iita.org/iws2016)

## 16 FEBRUARY

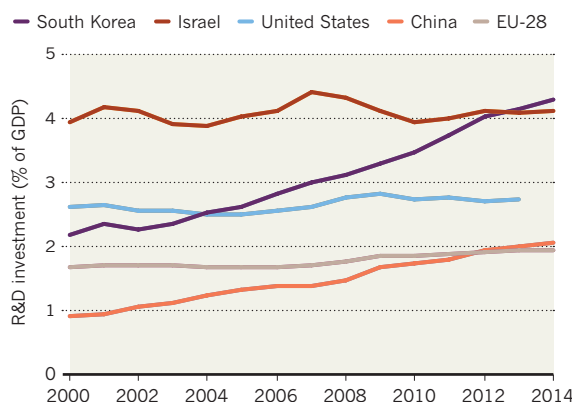
Sentinel-3A, a European Space Agency Earth-observing satellite, launches from Plesetsk, Russia. [go.nature.com/hgadp7](http://go.nature.com/hgadp7)

## TREND WATCH

South Korea is the world's most research-intensive economy, according to figures from the Organisation for Economic Co-operation and Development (OECD). In 2014, South Korea invested 4.29% of its gross domestic product (GDP) in research and development (R&D) — stretching ahead of Israel, which invests 4.11%. The South Korean increase is largely due to rising industrial investment, and the country is aiming to spend 5% of GDP on research by 2017.

## SOUTH KOREA'S RESEARCH BOOM

South Korean investment in research and development (R&D) is soaring as a percentage of its gross domestic product (GDP).



Rev. 1, CD005992; 2016). A group of systematics reviewers known as the Cochrane Collaboration assessed research from 21 countries and found improvements in cardiovascular health and reduced mortality from smoking-related illnesses since the first national ban, in Ireland in 2004. Effects on respiratory health, however, were “less consistent”. In total, 77 studies were assessed by the Cochrane team, and the findings strengthen the conclusions made in a previous review in 2010, the researchers say.

► [NATURE.COM](http://NATURE.COM)

For daily news updates see: [www.nature.com/news](http://www.nature.com/news)

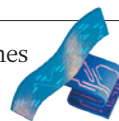


# NEWS IN FOCUS

**ETHICS** Sexual-harassment cases pose challenge for funders **p.138**

**SPACE** Astronomers hope it's third time lucky for Japanese probe **p.140**

**BIOTECH** Company publishes usually-secret replication failures **p.141**



**COMPUTING** Industry prepares to abandon pursuit of Moore's law **p.144**

DAN BROUIN



Charred pencil pines mark a fire-ravaged landscape in Tasmania, Australia.

AUSTRALIA

## Blazes threaten iconic trees

*As Tasmanian climate warms, bushfires are encroaching on forest ecosystems that date back more than 180 million years.*

BY EMMA MARRIS

**B**ushfires sparked by lightning storms have burned large areas of north-western Tasmania since mid-January, consuming more than 110,000 hectares — including almost 2% of the region's United Nations World Heritage wilderness lands.

Large Australian bushfires have become so common that they hardly seem like news, but these are different. Rather than burning swathes of fire-adapted vegetation such

as *Eucalyptus* or *Banksia*, many of the recent blazes have raged through ecosystems that are not adapted to fire at all — relict forests from when Tasmania was part of the Gondwana supercontinent, more than 180 million years ago.

When fires burn the pencil pines (*Athrotaxis cupressoides*) and King Billy pines (*Athrotaxis selaginoides*) that dominate these high-altitude ecosystems, they kill not only the trees but also their seeds — and incinerate the peaty soil in which they grow. In the

hardest-hit areas, natural regeneration of these ancient forests is all but impossible.

The trees, many of which are more than 1,000 years old, have been able to survive in Tasmania so far because its climate is colder and wetter than that of mainland Australia. But climate change may tip the odds against them.

The 'dry' lightning strikes — from thunderstorms in which most rain evaporates before reaching the ground — that ignited the current fires used to be vanishingly rare in western Tasmania, but climate change is thought to have ►

## FEELING THE BURN

Since mid-January, bushfires have blazed in northwestern Tasmania, threatening forest ecosystems that have not evolved to cope with such fires. Some of these landscapes date back more than 180 million years.

▨ Fires  
▨ World heritage site  
▨ Forest reserve  
▨ National park  
▨ State or regional reserve  
▨ Conservation area  
▨ Other designation

25 km

Indian Ocean



► made such storms more common in recent years. Tasmania saw one of the driest years on record in 2015. Furthermore, logging and dry conditions in the rainforest that surrounds these alpine forests have reduced its ability to act as a firebreak.

“There was no doubt pencil pine was on the mainland, but the fire and climate regime meant it couldn’t persist,” says David Lindenmayer, a forest ecologist at the Australian National University in Canberra. “If Tasmania is going to become more like the mainland, there is a distinct possibility that its time is going to be done. That is a huge loss for the world.”

Just a small percentage of the alpine and central-plateau fire-sensitive ecosystems in northwestern Tasmania have been killed by the blazes; 73 fires were still burning on 9 February.

But if fires become more frequent, that may spell the beginning of the end.

“If you have 2% [burning] every 10 or 15 years, it is not long before it is pushed into only the most fire-protected refuges,” says Jamie Kirkpatrick, a geographer and conservation ecologist at the University of Tasmania in Hobart.

To the chagrin of ecologists and bushwalkers, the fires have threatened several iconic landscapes, including the Walls of Jerusalem National Park and Mount Anne, the highest peak in southwestern Tasmania (see ‘Feeling the burn’). Hobart-based photographer Dan Broun and a colleague hiked up to a burned alpine area last weekend. When they crested the plateau, they were greeted by a dead world. “We were in shock. What we were seeing was

complete and utter devastation,” says Broun.

There have already been discussions about reseedling the burnt areas, but experts warn that it may not work; it is not at all clear that the strategy would make sense in a ‘new normal’ in which fires are predicted to run rampant in the area. Furthermore, grazing by wallabies and wombats keeps many seedlings from ever growing tall enough to reproduce. The existing pines may date from a slender window of time when heavy snows kept grazing animals away for long enough for the trees to grow out of their reach — a process that takes 50 years.

With threats to the region’s iconic forests clear, University of Tasmania ecologist David Bowman has called for an increased effort to collect seeds of fire-sensitive Tasmanian species and conserve them in another location, perhaps even on a sub-Antarctic island that would be safer from fire.

That would require a lot of seeds. Luckily, 2015 was a mast year — a period in which trees produce unusually large crops of cones and seeds — for many high-altitude conifers in Tasmania. In March and May, government biologists collected 1.57 million viable seeds from pencil and King Billy pines, as well as the shrub-like Cheshunt pine (*Diselma archeri*) and drooping pine (*Pherosphaera hookeriana*).

“We knew these sorts of [fire] events were likely in the future and the species we were looking at were so vulnerable to fire,” says James Wood, head of the Tasmanian Seed Conservation Centre at the Royal Tasmanian Botanical Gardens in Hobart. “Effectively, we are a Noah’s ark. We are just trying to hold onto material so it is not lost permanently.” ■

SOURCE: MAP GRANT WILLIAMSON/UNIV. TASMANIA; DATA: IUCN WORLD DATABASE OF PROTECTED AREAS/TASMANIAN FIRE SERVICE

## POLICY

# US science agencies grapple with sexual harassment

*Funders threaten harsh penalties, but many have yet to take action.*

BY ALEXANDRA WITZE

Fabricating, falsifying or plagiarizing data can get a grant yanked or a researcher blacklisted for breaking the professional code of science. Now, some funders are facing a fresh challenge: what to do with grants given to scientists who commit sexual transgressions.

The US government does not classify sexual infractions as research misconduct. Instead, as recent high-profile cases illustrate, the National Science Foundation (NSF), NASA and the National Institutes of Health (NIH)

must navigate a relatively new legal landscape when confronted with sexual harassment by grant recipients. What is clear, specialists in research ethics say, is that agencies, institutions and researchers all need to improve their response to such behaviour.

“The public has a right for us to conduct publicly funded work honourably and with integrity,” says C. K. Gunsalus, director of the National Center for Professional and Research Ethics at the University of Illinois at Urbana-Champaign.

The latest scandal broke on 2 February,

when *The New York Times* reported that molecular biologist Jason Lieb had left the University of Chicago in Illinois after an internal investigation found that he had made unwelcome sexual advances to female graduate students, and had engaged in sexual activity with a student who was incapacitated and could not consent. In January, it became public that Christian Ott, a theoretical astrophysicist at the California Institute of Technology in Pasadena, was on unpaid leave for violating the institution’s sexual-harassment policy. And last October, astronomer Geoffrey Marcy announced



that he would retire from the University of California, Berkeley, after a similar verdict.

Lieb, Ott and Marcy have each brought millions in government research dollars to their universities, which must now decide how to handle the money. Funding agencies and institutions must comply with Title IX, a federal law that forbids sex discrimination (which, in legal terms, can include harassment or assault) in any educational programme that receives government money. The law has been on the books since 1972, but in 2011, the US Department of Education said that it would step up its enforcement of the sexual-harassment aspects.

To comply with Title IX, institutions must have a representative who investigates and resolves allegations of sex discrimination. The inquiries into Lieb, Ott and Marcy all went through the Title IX offices at their respective universities.

Once a Title IX investigation is complete, the institution must decide whether to take disciplinary action. A funding agency can open its own inquiry and levy extra penalties if it deems them necessary. NASA and the NSF have both put out statements recently saying that they



The Lick Observatory in California is home to an exoplanet search once run by Geoffrey Marcy, who left the University of California, Berkeley, after a sexual-harassment investigation.

do not tolerate sexual harassment by grantees; the NSF even threatened to pull funds entirely from institutions that do not comply with Title IX. But it has never banned a grantee, let alone an institution, for violating Title IX.

"Statements are good for changing the culture, but they have to be supported by action," says Katie Hinde, an anthropologist at Arizona State University in Tempe.

The process is more complicated than simply yanking existing grants. "People come to us and say, why don't you fix it?" says James Ulvestad, head of the NSF's astronomy division. "Well, what the funding agencies can do is what federal law allows us to do."

In general, funding awards are made to institutions, not the person who is the principal investigator (PI) for the work. Even if a PI has been found to violate institutional policies, his or her grant money will continue to flow to support graduate students, postdoctoral fellows and other collaborators on an affected project.

Lieb is PI or co-PI on more than US\$1.2 million in NIH grants, and the University of Chicago may nominate someone to replace him. Ott is involved in more than \$3.2 million in NSF funds; those grants are under review but their status remains unchanged for now, says NSF spokeswoman Ivy Kupec. Ott still has access to NSF facilities, including the Blue Waters supercomputer in Illinois, on which he runs simulations of supernova explosions.

Marcy had many lucrative private grants, including a share of the \$100-million Breakthrough Listen search for extraterrestrial intelligence. Those have been transferred to new PIs.

His old employer is now working on designating new PIs for the two NASA grants totalling roughly \$1 million that are in Marcy's name.

It was actually a case of sexual assault that prompted the modern US definition of research misconduct, says Nicholas Steneck, a specialist in research integrity at the University of Michigan in Ann Arbor. In 1989, the NSF received complaints about the actions of a senior researcher while taking undergraduates to a research site in another country. The agency's inspector-general ultimately found the researcher to have committed "16 incidents of sexual misfeasance with female graduate and undergraduate students at the research site; on the way to the site; and in his home, car, and office". Many could be classified as sexual assaults.

The agency barred him from receiving federal funding for five years, according to a definition of misconduct that at that time included "serious deviations" from accepted research practices. After a fierce battle over the meaning of the phrase, agencies excised it from their definitions of misconduct in 2000. "Sexual harassment shouldn't be tolerated," Steneck says. "But it's not so easy to say, 'It's research misconduct and that's the way it ought to be handled'."

US agencies might learn a thing or two from funders in other countries, Steneck adds. Both Canada and Australia (see 'Codes of conduct') require federally funded scientists to meet a minimum ethical standard that specifically describes institutional roles.

"You need to set out a clear code of what to expect," he says. ■

## CODES OF CONDUCT

Countries have very different laws governing how to deal with sexual misconduct in recipients of federal research grants. Some link it closely to research misconduct; others address the issue as more of a workplace violation.

### CANADA

A three-agency framework for responsible conduct of research says that "researchers shall follow the requirements of applicable institutional policies and professional or disciplinary standards".

### AUSTRALIA

The Australian Code for the Responsible Conduct of Research classifies "other forms of misconduct, such as harassment, bullying or financial misconduct" as covered by employee relationships with their institutions.

### EUROPE

The European Code of Conduct for Research Integrity includes, among its principles of integrity, the "responsibility for the scientists and researchers of the future".

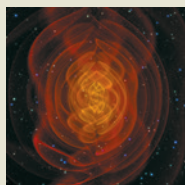
### UNITED STATES

A phrase that covers sexual infractions was removed from the definition of research misconduct in 2000. But funders must adhere to a law known as Title IX, which prohibits discrimination on the basis of sex.



  
**MORE  
ONLINE**

### TOP STORY



The latest in physicists' hunt for gravitational waves  
[go.nature.com/jnqkfd](http://go.nature.com/jnqkfd)

### MORE NEWS

- Nobel official resigns over Karolinska surgeon controversy  
[go.nature.com/rh1ne4](http://go.nature.com/rh1ne4)
- Microplastics damage oyster fertility  
[go.nature.com/vwnijx](http://go.nature.com/vwnijx)
- Destroying worn-out cells makes mice live longer  
[go.nature.com/vqhegg](http://go.nature.com/vqhegg)

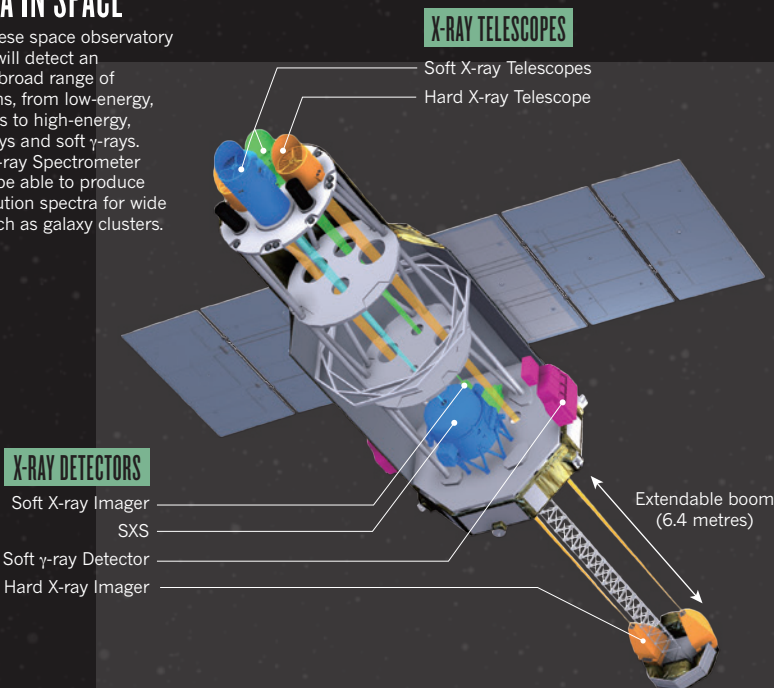
### NATURE PODCAST



The end of Moore's law; religion and cooperation; and stock-market climate risks  
[nature.com/nature/podcast](http://nature.com/nature/podcast)

## SPECTRA IN SPACE

The Japanese space observatory ASTRO-H will detect an unusually broad range of wavelengths, from low-energy, 'soft' X-rays to high-energy, 'hard' X-rays and soft  $\gamma$ -rays. The Soft X-ray Spectrometer (SXS) will be able to produce high-resolution spectra for wide objects such as galaxy clusters.



LENGTH Approx. 14 m | MASS 2.7 tonnes | POWER 3,500 watts | LIFE 3 years

is also a member of ASTRO-H's ESA team.

Spectra carry information about the velocity and turbulence of the plasma that pervades galaxy clusters, which in turn can reveal whether, and how, a cluster resulted from a merger of two smaller ones, says Christine Forman, a high-energy astrophysicist at Harvard University in Cambridge, Massachusetts. X-ray pictures of clusters, when combined with visible-light images, have also previously provided striking — although indirect — evidence for the existence of dark matter. ASTRO-H should be able to help settle whether a 3.5-kilo-electronvolt X-ray signal seen in certain galaxies is a signature of dark matter decaying into photons — or something else, says Alexey Boyarsky, an astrophysicist at Leiden University in the Netherlands who co-discovered the signal (see *Nature* 517, 422–423; 2015).

ASTRO-H will also cover a broader range of wavelengths than most other missions, from 'soft', low-energy photons starting at 300 eV, through hard X-rays and right up to soft  $\gamma$ -rays of 600 keV. But only the soft X-rays will have their spectra imaged to high resolution. At the heart of the instrument responsible for this — the Soft X-ray Spectrometer (SXS) — is an array of 36-pixel elements that must be kept at 0.05 degrees above absolute zero, well inside the main body of the craft. When a photon strikes one of the sensors, the temperature of the sensor rises slightly, causing its electrical conductivity to increase. The resulting change in voltage can be used to measure the energy of the original photon — and therefore its wavelength — with a precision of nearly one part in 1,000.

The technology first took shape in 1984, when Richard Kelley, an astrophysicist at NASA's Goddard Space Flight Center in Greenbelt, Maryland, started designing sensors for Chandra. After NASA scaled back its ambitions for the observatory and scrapped that plan, Kelley and his team provided the spectrometer for a Japanese X-ray telescope called ASTRO-E. But in 2000, the rocket that was supposed to put it into orbit crashed shortly after take-off.

JAXA prepared a replacement mission, Suzaku, which reached orbit in July 2005 — only for disaster to strike again. The liquid helium used to keep a spectrometer's sensors cold has to be slowly released, or outgassed. But in Suzaku, a tiny amount of that helium accumulated inside the craft. This was enough to ruin the vacuum that was supposed to insulate the helium tank from the rest of the craft. The tank heated up faster than expected, causing the helium to boil off and vent into space within four weeks of launch. Unable to stay super-chilled, the spectrometer was crippled before it could start making observations, although Suzaku's other instruments continued to operate until the craft was decommissioned in 2015.

For ASTRO-H, JAXA redesigned the tank with plumbing to outgas the helium straight into space. And to further guard against glitches,

SOURCE: JAXA

## ASTRONOMY

# High stakes for Japan's space probe

*ASTRO-H will carry a technology that two earlier, ill-fated space telescopes failed to put into action.*

BY DAVIDE CASTELVECCHI

Astronomers around the world are hoping that the third time will be the charm as Japan prepares to launch its largest space observatory ever. The telescope will use X-rays to study phenomena from black holes to dark matter and carries cryogenic imaging technology that flew on two previous missions — but met disaster both times.

Weather permitting, a Japanese-built H-IIA rocket is scheduled to launch the probe, provisionally named ASTRO-H, from Tanegashima Space Center at 17:45 local time on 12 February. Once in orbit, the 2.7-tonne probe will stretch out to its maximum length of 14 metres, including a 6.4-metre boom that will host an imager capable of collecting high-energy, or 'hard' X-ray photons (see 'Spectra in space'). The Japan Aerospace Exploration Agency (JAXA) leads the mission with an investment of ¥31 billion (US\$265 million), but there is

also major participation from NASA, as well as institutions in six other nations and the European Space Agency (ESA).

Studying X-ray emissions is the best way to observe a wide range of cosmic phenomena, from galaxy clusters to the super-heated accretion disks around black holes. But Earth's atmosphere is mostly opaque to radiation outside the visible spectrum, and particularly to X-rays and  $\gamma$ -rays, meaning that most X-ray astronomy requires a satellite.

The major existing X-ray satellites are NASA's Chandra X-ray Observatory and ESA's XMM-Newton, which both launched in 1999. These can analyse the constituent wavelengths of X-rays — the spectra — emitted by point-like objects such as stars. But ASTRO-H will be the first to provide high-resolution spectra for much more spread-out X-ray sources such as galaxy clusters, says Norbert Scharrel, project scientist for XMM-Newton at ESA's European Space Astronomy Centre outside Madrid, who



ASTRO-H project manager Tadayuki Takahashi, an astrophysicist at the University of Tokyo and ASTRO-H's project manager, pushed his collaboration to work without borders. "Usually, international coalitions have clearly defined interfaces," with different laboratories providing isolated components of a spacecraft and its payload, says Takahashi. But ASTRO-H researchers regularly visited each other's labs, sometimes for months at a time.

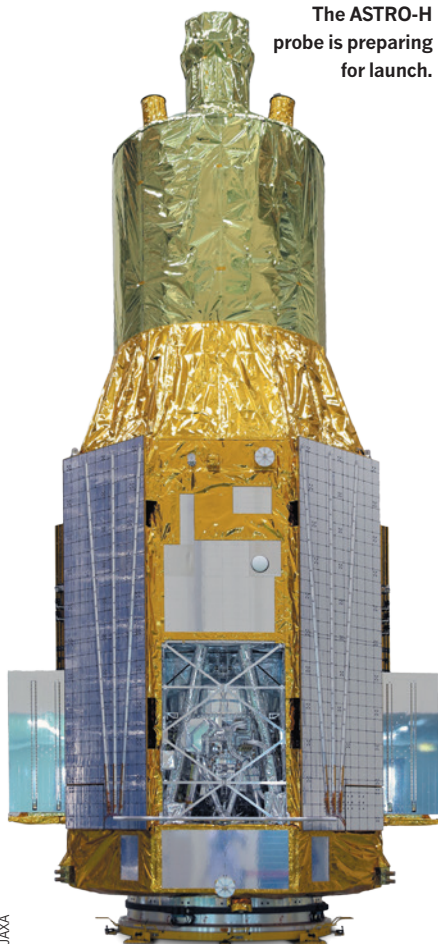
Kelley says that Takahashi forged a very open collaboration. "Tad understands that if you want to maximize the chances of success, you have to have no barriers," he says. "Everybody has access to everything."

Astronomers around the world will be allowed to request observing time with ASTRO-H. Each team will have exclusive access to the resulting data for one year, after which JAXA will make them publicly available — a model long adopted by NASA. ASTRO-H will be renamed after launch, although what it will be called is yet to be determined.

A larger, higher-resolution version of the SXS is due to fly aboard Athena, an ESA-led X-ray astronomy mission planned for the late 2020s. ■

Additional reporting by David Cyranoski.

The ASTRO-H probe is preparing for launch.



JAXA

## REPLICATION

# Biotech giant posts negative results

*Amgen papers seed channel for discussing reproducibility.*

BY MONYA BAKER

A biotechnology firm is releasing data on three failed efforts to confirm findings published in high-profile scientific journals — details that the industry usually keeps secret.

Amgen, headquartered in Thousand Oaks, California, says that it hopes the move will encourage others to describe their own replication attempts, and thus help the scientific community to get to the bottom of work that other labs are having trouble verifying.

The data are posted online on a newly launched channel dedicated to quickly publishing researchers' efforts to confirm scientific findings (see [go.nature.com/3zzea9](http://go.nature.com/3zzea9)), hosted by *F1000Research*, the publishing platform of London-based publishers Faculty of 1000 (F1000).

The idea emerged from discussions at a meeting in 2015 focused on improving scientific integrity. Sasha Kamb, who leads research discovery at Amgen, said that his company's scientists have many times failed to reproduce academic studies but that it takes too much effort to publish these accounts.

Bruce Alberts, a former editor-in-chief of *Science* who sits on *F1000Research*'s advisory board, suggested that Kamb try the faster F1000 route — an open-science publishing model in which submitted studies are posted online (for a fee that ranges from US\$150 to \$1,000) before undergoing open peer review; submissions are subject to checks by F1000 editors to ensure that data are freely available and that methods are adequately described. "The idea is to get the data out and get it critically looked at," Alberts says.

## IRREPRODUCIBLE HISTORY

In 2012, Amgen researchers declared that they had been unable to reproduce the findings in 47 of 53 'landmark' cancer papers<sup>1</sup>. Those papers were never identified — partly because of confidentiality concerns — and there are no plans to release details now, says Kamb, who was not involved with that work.

The three latest studies that Amgen has posted deliberately do not make a detailed comparison of their results to previous papers, says Kamb. "We don't want to

make strong conclusions that someone else's work is wrong with a capital W," he says.

One study adds to existing criticism of a *Science* paper<sup>2</sup> that suggested a cancer drug might be a potential treatment for Alzheimer's disease; a second counters earlier findings (including some by Amgen researchers) connecting a gene to insulin sensitivity in mice<sup>3,4</sup>; and a third counters a *Nature* paper<sup>5</sup> reporting that inhibiting one particular protein could enhance the degradation of others associated with neurodegenerative diseases. Amgen researchers did not contact the original authors when they conducted their studies, Kamb says.

## GAINING TRACTION

The main way that the scientific community spreads the word about irreproducible research is through insinuation, which is inefficient and unfair to the original researchers, says Ricardo Dolmetsch, global head of neuroscience at Novartis's Institutes for Biomedical Research in Cambridge, Massachusetts. "Anything we can do to improve the ratio of signal to noise in the literature is very welcome," he says.

The F1000 initiative is useful, but previous efforts have tried and failed to encourage the reporting of replications and negative results, cautions John Ioannidis, who studies scientific robustness at California's Stanford University. In general, the scientific community undervalues such work, he says.

But Kamb says that he has spoken with several industry leaders who have expressed support, and he hopes that they will contribute eventually. Morgan Sheng, a vice-president at biotechnology company Genentech in South San Francisco, says that he can foresee his company's scientists submitting data to the venture. "I believe the main risk of a publication venue like the F1000 channel is that it becomes a place for 'bashing' good science, because biological experiments are complex and beset by many variables that are hard to control. Non-replication does not necessarily mean 'not true,'" Sheng adds. ■

1. Begley, C. G. & Ellis, L. M. *Nature* **483**, 531–533 (2012).
2. Cramer, P. E. *et al. Science* **335**, 1503–1506 (2012).
3. Gardner, J. *et al. Biochem. Biophys. Res. Commun.* **418**, 1–5 (2012).
4. Osborn, O. *et al. J. Clin. Invest.* **122**, 2444–2453 (2012).
5. Lee, B.-H. *et al. Nature* **467**, 179–184 (2010).

**NATURE.COM**  
For a longer version  
of this story, see  
[go.nature.com/qghfju](http://go.nature.com/qghfju)

## GENETICS

# US panel backs 'three-person' embryos

*But regulators cannot approve clinical trials.*

BY SARA REARDON

The US Food and Drug Administration (FDA) should approve clinical trials of a gene-therapy technique to create embryos with genetic material from three people — if certain conditions are met, the US National Academies of Sciences, Engineering and Medicine said on 3 February.

The controversial technique involves replacing the faulty energy-producing mitochondria in a woman's egg with mitochondria from the egg of a healthy donor. The aim is to prevent the transmission of disabling or fatal diseases caused by mutations in a mother's mitochondrial DNA. But concerns about the technique's safety, and the psychological and social implications of children with three genetic parents, have given US regulators pause. Moreover, a federal law approved late last year prevents the FDA from allowing any such trials in humans.

The academy panel suggests limiting the tests of mitochondrial replacement to male embryos as a safety precaution. Male offspring would not be able to pass their modified mitochondria to future generations, because a child inherits its mitochondria from its mother.

The report also outlines several extra steps to monitor the safety of mitochondrial replacement, such as attempting to track the children born as a result of the technique and sharing the resulting data with scientists and the public. If mitochondrial replacement is proved safe in male offspring, it could be expanded to female embryos, the advisory panel said.

By contrast, the United Kingdom last year approved mitochondrial replacement with no restrictions on the sex of a modified embryo.

Shoukhrat Mitalipov, a reproductive-biology specialist at the Oregon Health and Science University in Portland, sees the US advisory panel report as a hollow victory. In 2014, he applied to the FDA to perform a clinical trial of mitochondrial replacement therapy with human embryos — a trial that cannot take place under current US law.

Others disagree with the panel's conclusions. The new report "leapt to the conclusion that things should go forward despite all the concerns", says Marcy Darnovsky, executive director of the non-profit Center for Genetics and Society in Berkeley, California. ■



Pregnant women infected with the Zika virus are being monitored for effects on their unborn babies.

## PUBLIC HEALTH

# Scientists probe Zika link to birth defects

*Mosquito-borne virus is now present in 33 countries.*

BY ERIKA CHECK HAYDEN

Public-health authorities are investigating whether the Zika virus has caused an apparent surge in the number of infants born with microcephaly, or abnormally small heads, in at least seven countries. But conclusively determining whether the mosquito-borne virus is to blame could take months to years, researchers say.

Concerns first arose in Brazil, which in November declared a national public-health emergency. As of 2 February, officials there had investigated 1,113 of 4,783 suspected cases of microcephaly reported since late last year, and confirmed that 404 of those could be linked to Zika. On 1 February, a committee convened by the World Health Organization said that a causal link between Zika and microcephaly is "strongly suspected, though not yet scientifically proven".

That is not for a lack of effort. Work now under way includes case-control studies to compare rates of Zika infection in babies who are born with microcephaly and in those without it, as well as genetic sequencing of

the virus and efforts to develop a molecular diagnostic test for Zika infection.

Making progress has been difficult because scientists know relatively little about Zika; there is no easy-to-use test to diagnose infections; and physicians disagree about how to define microcephaly, says Bruno Andrade, an immunologist at the Fiocruz research institute in Bahia, Brazil. "All of this started less than two months ago — that's when everything stepped up," he says. "We are in the middle of this nightmare here."

So far, two lines of evidence support a link between the virus and microcephaly. Microcephaly cases in Brazil started to rise around 6 months after authorities confirmed Zika transmission there, hinting that the defect might have been caused by *in utero* exposure to the virus. And researchers in Brazil have found traces of the virus, or antibodies to it, in the amniotic fluid, brains or spinal fluid of 15 fetuses and babies diagnosed with microcephaly.

This is suggestive, but not conclusive. "Most of us believe it's highly plausible that Zika is the cause of this epidemic of microcephaly,

SCHNEIDER MENDOZA/AFP/GETTY



but we need additional evidence,” says Albert Ko, an infectious-disease physician and epidemiologist at the Yale School of Public Health in New Haven, Connecticut.

In the hope of producing more-definitive data, the Brazilian ministry of health is now setting up large studies. In one, researchers will follow up 6,000 pregnant women in northeastern Brazil to investigate the effects of Zika and of microcephaly.

#### KNOWLEDGE GAPS

Epidemiological studies are often complex because Zika causes a relatively mild illness in adults and there is no widely deployed test for the virus. This means that most of the mothers who have participated in previous studies were never diagnosed with Zika, even if they had it. To address this problem, the Brazilian health ministry is now asking mothers whether they had Zika symptoms instead of whether they were diagnosed with Zika.

But many researchers say that epidemiological data alone will not convince them of a link between Zika and microcephaly; they would like to see evidence of how and why the virus causes the condition. With this in mind, scientists are developing animal models to investigate Zika's effects on the body, such as which tissues it infects, and why fetal brains might be especially vulnerable.

“There’s a lot of basic work and research that we need to do,” says Anthony Fauci, director of the US National Institute of Allergy and Infectious Diseases in Bethesda, Maryland.

It’s plausible, for instance, that Zika crosses from mother to baby through the placenta, as do some related viruses, such as West Nile virus. But these other viruses don’t often cause infant brain damage, so it’s not clear why — or

**“There’s a lot of basic work and research that we need to do.”**

how — Zika might, says David Morens, a senior adviser to Fauci.

The virus may be toxic only while a fetus’s brain is still developing its major structures, in the first two months of pregnancy. Or it may persist in the body for a long period, which would explain why Zika is seen in still-born babies with microcephaly. “If the insult happened early on, then why is the virus present at seven months when the miscarriage occurs?” Morens says. “There must be a combination of things going on.”

Another conundrum is what it is that makes these women and babies so vulnerable: the vast majority of women infected with Zika go on to have healthy babies.

But whatever scientists ultimately discover about Zika, another major challenge remains: supporting the children

born with microcephaly. Many are living in what Ko calls a “vegetative state”, and may go on to have seizures. That could be difficult for their families — many of which are impoverished — to bear.

“We have to start settling down and think about how we’re going to take care of these kids,” Ko says. If there is a link between Zika and microcephaly, the number of babies affected by the condition could soar as the virus spreads, he cautions. “We have no idea how big it’s going to get.” ■

#### CORRECTIONS

The two elements in the key for the map in the News story ‘Microcephaly surge in doubt’ (*Nature* **530**, 13–14; 2016) were accidentally swapped round. The corrected graphic can be seen at [go.nature.com/ekmksk](http://go.nature.com/ekmksk). The dollar conversion of €4.6 billion in the News story ‘Science hubs win in major revamp’ (*Nature* **530**, 18–19; 2016) should have been US\$5 billion not \$5 million. The News story ‘Academics across Europe join ‘Brexit’ debate’ (*Nature* **530**, 15; 2016) erroneously stated that the latest UK contribution to the EU budget was €140 billion. In fact, it contributed around €17 billion.



# MORE THAN MOORE

A surreal illustration of a person standing on a circuit board, surrounded by floating and tilted circuit boards and components against a dark blue background with stars.

BY M. MITCHELL WALDROP



# THE SEMICONDUCTOR INDUSTRY WILL SOON ABANDON ITS PURSUIT OF MOORE'S LAW. NOW THINGS COULD GET A LOT MORE INTERESTING.

**N**ext month, the worldwide semiconductor industry will formally acknowledge what has become increasingly obvious to everyone involved: Moore's law, the principle that has powered the information-technology revolution since the 1960s, is nearing its end.

A rule of thumb that has come to dominate computing, Moore's law states that the number of transistors on a microprocessor chip will double every two years or so — which has generally meant that the chip's performance will, too. The exponential improvement that the law describes transformed the first crude home computers of the 1970s into the sophisticated machines of the 1980s and 1990s, and from there gave rise to high-speed Internet, smartphones and the wired-up cars, refrigerators and thermostats that are becoming prevalent today.

None of this was inevitable: chipmakers deliberately chose to stay on the Moore's law track. At every stage, software developers came up with applications that strained the capabilities of existing chips; consumers asked more of their devices; and manufacturers rushed to meet that demand with next-generation chips. Since the 1990s, in fact, the semiconductor industry has released a research road map every two years to coordinate what its hundreds of manufacturers and suppliers are doing to stay in step with the law — a strategy sometimes called More Moore. It has been largely thanks to this road map that computers have followed the law's exponential demands.

Not for much longer. The doubling has already started to falter, thanks to the heat that is unavoidably generated when more and more silicon circuitry is jammed into the same small area. And some even more fundamental limits loom less than a decade away. Top-of-the-line microprocessors currently have circuit features that are around 14 nanometres across, smaller than most viruses. But by the early 2020s, says Paolo Gargini, chair of the road-mapping organization, “even with super-aggressive efforts, we'll get to the 2–3-nanometre limit, where features are just 10 atoms across. Is that a device at all?” Probably not — if only because at that scale, electron behaviour will be governed by quantum uncertainties that will make transistors hopelessly unreliable. And despite vigorous research efforts, there is no obvious successor to today's silicon technology.

The industry road map released next month will for the first time lay out a research and development plan that is not centred on Moore's law. Instead, it will follow what might be called the More than Moore strategy: rather than making the chips better and letting the applications follow, it will start with applications — from smartphones and supercomputers to data centres in the cloud — and work downwards to see what chips are needed to support them. Among those chips will be new generations of sensors, power-management circuits and other silicon devices required by a world in which computing is increasingly mobile.

The changing landscape, in turn, could splinter the industry's long tradition of unity in pursuit of Moore's law. “Everybody is struggling with what the road map actually means,” says Daniel Reed, a computer scientist and vice-president for research at the University of Iowa in Iowa

City. The Semiconductor Industry Association (SIA) in Washington DC, which represents all the major US firms, has already said that it will cease its participation in the road-mapping effort once the report is out, and will instead pursue its own research and development agenda.

Everyone agrees that the twilight of Moore's law will not mean the end of progress. “Think about what happened to airplanes,” says Reed. “A Boeing 787 doesn't go any faster than a 707 did in the 1950s — but they are very different airplanes”, with innovations ranging from fully electronic controls to a carbon-fibre fuselage. That's what will happen with computers, he says: “Innovation will absolutely continue — but it will be more nuanced and complicated.”

## LAYING DOWN THE LAW

The 1965 essay<sup>1</sup> that would make Gordon Moore famous started with a meditation on what could be done with the still-new technology of integrated circuits. Moore, who was then research director of Fairchild Semiconductor in San Jose, California, predicted wonders such as home computers, digital wristwatches, automatic cars and “personal portable communications equipment” — mobile phones. But the heart of the essay was Moore's attempt to provide a timeline for this future. As a measure of a microprocessor's computational power, he looked at transistors, the on-off switches that make computing digital. On the basis of achievements by his company and others in the previous few years, he estimated that the number of transistors and other electronic components per chip was doubling every year.

Moore, who would later co-found Intel in Santa Clara, California, underestimated the doubling time; in 1975, he revised it to a more realistic two years<sup>2</sup>. But his vision was spot on. The future that he predicted started to arrive in the 1970s and 1980s, with the advent of microprocessor-equipped consumer products such as the Hewlett Packard hand calculators, the Apple II computer and the IBM PC. Demand for such products was soon exploding, and manufacturers were engaging in a brisk competition to offer more and more capable chips in smaller and smaller packages (see ‘Moore's lore’).

This was expensive. Improving a microprocessor's performance meant scaling down the elements of its circuit so that more of them could be packed together on the chip, and electrons could move between them more quickly. Scaling, in turn, required major refinements in photolithography, the basic technology for etching those microscopic elements onto a silicon surface. But the boom times were such that this hardly mattered: a self-reinforcing cycle set in. Chips were so versatile that manufacturers could make only a few types — processors and memory, mostly — and sell them in huge quantities. That gave them enough cash to cover the cost of upgrading their fabrication facilities, or ‘fabs’, and still drop the prices, thereby fuelling demand even further.

Soon, however, it became clear that this market-driven cycle could not sustain the relentless cadence of Moore's law by itself. The chip-making process was getting too complex, often involving hundreds of stages, which meant that taking the next step down in scale required a network of materials-suppliers and apparatus-makers to deliver the right upgrades at the right time. “If you need 40 kinds of equipment and only 39 are ready, then everything stops,” says Kenneth Flamm, an economist who studies the computer industry at the University of Texas at Austin.

To provide that coordination, the industry devised its first road map. The idea, says Gargini, was “that everyone would have a rough estimate of where they were going, and they could raise an alarm if they saw roadblocks ahead”. The US semiconductor industry launched the mapping effort in 1991, with hundreds of engineers from various companies working on the first report and its subsequent iterations, and Gargini, then the director of technology strategy at Intel, as its chair. In 1998,

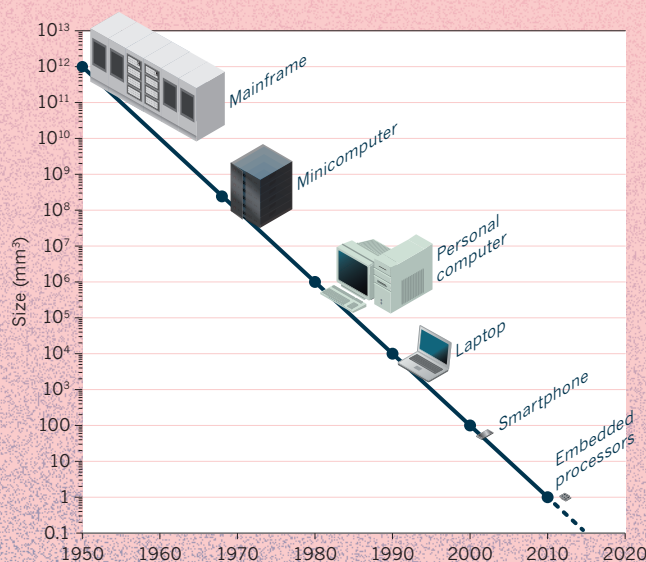
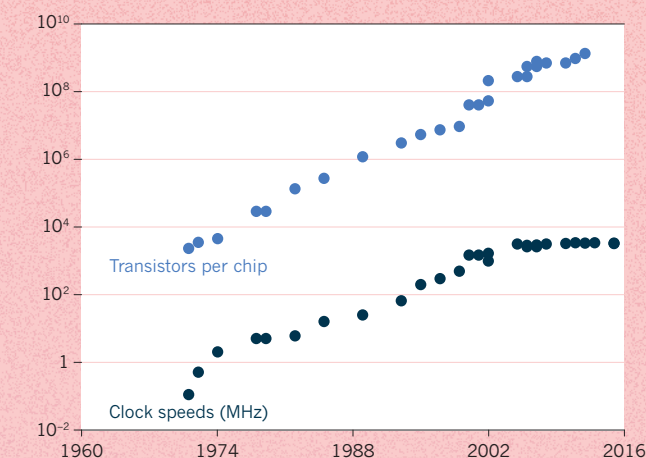
the effort became the International Technology Roadmap for Semiconductors, with participation from industry associations in Europe, Japan, Taiwan and South Korea. (This year's report, in keeping with its new approach, will be called the International Roadmap for Devices and Systems.)

## ➔ NATURE.COM

To hear more about what will come after Moore's law, visit: [go.nature.com/nppjyx](http://go.nature.com/nppjyx)

## MOORE'S LORE

For the past five decades, the number of transistors per microprocessor chip — a rough measure of processing power — has doubled about every two years, in step with Moore's law (top). Chips also increased their 'clock speed', or rate of executing instructions, until 2004, when speeds were capped to limit heat. As computers increase in power and shrink in size, a new class of machines has emerged roughly every ten years (bottom).



"The road map was an incredibly interesting experiment," says Flamm. "So far as I know, there is no example of anything like this in any other industry, where every manufacturer and supplier gets together and figures out what they are going to do." In effect, it converted Moore's law from an empirical observation into a self-fulfilling prophecy: new chips followed the law because the industry made sure that they did.

And it all worked beautifully, says Flamm — right up until it didn't.

### HEAT DEATH

The first stumbling block was not unexpected. Gargini and others had warned about it as far back as 1989. But it hit hard nonetheless: things got too small.

"It used to be that whenever we would scale to smaller feature size, good things happened automatically," says Bill Bottoms, president of Third Millennium Test Solutions, an equipment manufacturer in Santa Clara. "The chips would go faster and consume less power."

But in the early 2000s, when the features began to shrink below about 90 nanometres, that automatic benefit began to fail. As electrons had to move faster and faster through silicon circuits that were smaller and smaller, the chips began to get too hot.

That was a fundamental problem. Heat is hard to get rid of, and no one wants to buy a mobile phone that burns their hand. So manufacturers seized on the only solutions they had, says Gargini. First, they stopped trying to increase 'clock rates' — how fast microprocessors execute instructions. This effectively put a speed limit on the chip's electrons and limited their ability to generate heat. The maximum clock rate hasn't budged since 2004.

Second, to keep the chips moving along the Moore's law performance curve despite the speed limit, they redesigned the internal circuitry so that each chip contained not one processor, or 'core', but two, four or more. (Four and eight are common in today's desktop computers and smartphones.) In principle, says Gargini, "you can have the same output with four cores going at 250 megahertz as one going at 1 gigahertz". In practice, exploiting eight processors means that a problem has to be broken down into eight pieces — which for many algorithms is difficult to impossible. "The piece that can't be parallelized will limit your improvement," says Gargini.

Even so, when combined with creative redesigns to compensate for electron leakage and other effects, these two solutions have enabled chip manufacturers to continue shrinking their circuits and keeping their transistor counts on track with Moore's law. The question now is what will happen in the early 2020s, when continued scaling is no longer possible with silicon because quantum effects have come into play. What comes next? "We're still struggling," says An Chen, an electrical engineer who works for the international chipmaker GlobalFoundries in Santa Clara, California, and who chairs a committee of the new road map that is looking into the question.

That is not for a lack of ideas. One possibility is to embrace a completely new paradigm — something like quantum computing, which promises exponential speed-up for certain calculations, or neuromorphic computing, which aims to model processing elements on neurons in the brain. But none of these alternative paradigms has made it very far out of the laboratory. And many researchers think that quantum computing will offer advantages only for niche applications, rather than for the everyday tasks at which digital computing excels. "What does it mean to quantum-balance a chequebook?" wonders John Shalf, head of computer-science research at the Lawrence Berkeley National Laboratory in Berkeley, California.

### MATERIAL DIFFERENCES

A different approach, which does stay in the digital realm, is the quest to find a 'millivolt switch': a material that could be used for devices at least as fast as their silicon counterparts, but that would generate much less heat. There are many candidates, ranging from 2D graphene-like compounds to spintronic materials that would compute by flipping electron spins rather than by moving electrons. "There is an enormous research space to be explored once you step outside the confines of the established technology," says Thomas Theis, a physicist who directs the nanoelectronics initiative at the Semiconductor Research Corporation (SRC), a research-funding consortium in Durham, North Carolina.

Unfortunately, no millivolt switch has made it out of the laboratory either. That leaves the architectural approach: stick with silicon, but configure it in entirely new ways. One popular option is to go 3D. Instead of etching flat circuits onto the surface of a silicon wafer, build skyscrapers: stack many thin layers of silicon with microcircuitry etched into each. In principle, this should make it possible to pack more computational power into the same space. In practice, however, this currently works only with memory chips, which do not have a heat problem: they use circuits that consume power only when a memory cell is accessed, which is not that often. One example is the Hybrid Memory Cube design, a stack of as many as eight memory layers that is being pursued by an industry consortium originally

SOURCE: TOP: INTEL; BOTTOM: SIA/SRC



launched by Samsung and memory-maker Micron Technology in Boise, Idaho.

Microprocessors are more challenging: stacking layer after layer of hot things simply makes them hotter. But one way to get around that problem is to do away with separate memory and microprocessing chips, as well as the prodigious amount of heat — at least 50% of the total — that is now generated in shuttling data back and forth between the two. Instead, integrate them in the same nanoscale high-rise.

This is tricky, not least because current-generation microprocessors and memory chips are so different that they cannot be made on the same fab line; stacking them requires a complete redesign of the chip's structure. But several research groups are hoping to pull it off. Electrical engineer Subhashish Mitra and his colleagues at Stanford University in California have developed a hybrid architecture that stacks memory units together with transistors made from carbon nanotubes, which also carry current from layer to layer<sup>3</sup>. The group thinks that its architecture could reduce energy use to less than one-thousandth that of standard chips.

## GOING MOBILE

The second stumbling block for Moore's law was more of a surprise, but unfolded at roughly the same time as the first: computing went mobile.

Twenty-five years ago, computing was defined by the needs of desktop and laptop machines; supercomputers and data centres used essentially the same microprocessors, just packed together in much greater numbers. Not any more. Today, computing is increasingly defined by what high-end smartphones and tablets do — not to mention by smart watches and other wearables, as well as by the exploding number of smart devices in everything from bridges to the human body. And these mobile devices have priorities very different from those of their more sedentary cousins.

Keeping abreast of Moore's law is fairly far down on the list — if only because mobile applications and data have largely migrated to the worldwide network of server farms known as the cloud. Those server farms now dominate the market for powerful, cutting-edge microprocessors that do follow Moore's law. "What Google and Amazon decide to buy has a huge influence on what Intel decides to do," says Reed.

Much more crucial for mobiles is the ability to survive for long periods on battery power while interacting with their surroundings and users. The chips in a typical smartphone must send and receive signals for voice calls, Wi-Fi, Bluetooth and the Global Positioning System, while also sensing touch, proximity, acceleration, magnetic fields — even fingerprints. On top of that, the device must host special-purpose circuits for power management, to keep all those functions from draining the battery.

The problem for chipmakers is that this specialization is undermining the self-reinforcing economic cycle that once kept Moore's law humming. "The old market was that you would make a few different things, but sell a whole lot of them," says Reed. "The new market is that you have to make a lot of things, but sell a few hundred thousand apiece — so it had better be really cheap to design and fab them."

Both are ongoing challenges. Getting separately manufactured technologies to work together harmoniously in a single device is often a nightmare, says Bottoms, who heads the new road map's committee on the subject. "Different components, different materials, electronics, photonics and so on, all in the same package — these are issues that will have to be solved by new architectures, new simulations, new switches and more."

For many of the special-purpose circuits, design is still something of a cottage industry — which means slow and costly. At the University of California, Berkeley, electrical engineer Alberto Sangiovanni-Vincentelli and his colleagues are trying to change that: instead of starting from scratch each time, they think that people should create new devices by combining large chunks of existing circuitry that have known functionality<sup>4</sup>. "It's like using Lego blocks," says Sangiovanni-Vincentelli. It's a challenge to make sure that the blocks work together, but "if you were to use older methods of design, costs would be prohibitive".

Costs, not surprisingly, are very much on the chipmakers' minds these days. "The end of Moore's law is not a technical issue, it is an economic

issue," says Bottoms. Some companies, notably Intel, are still trying to shrink components before they hit the wall imposed by quantum effects, he says. But "the more we shrink, the more it costs".

Every time the scale is halved, manufacturers need a whole new generation of ever more precise photolithography machines. Building a new fab line today requires an investment typically measured in many billions of dollars — something only a handful of companies can afford. And the fragmentation of the market triggered by mobile devices is making it harder to recoup that money. "As soon as the cost per transistor at the next node exceeds the existing cost," says Bottoms, "the scaling stops."

Many observers think that the industry is perilously close to that point already. "My bet is that we run out of money before we run out of physics," says Reed.

Certainly it is true that rising costs over the past decade have forced a massive consolidation in the chip-making industry. Most of the world's production lines now belong to a comparative handful of multinationals such as Intel, Samsung and the Taiwan Semiconductor Manufacturing Company in Hsinchu. These manufacturing giants have tight relationships with the companies that supply them with materials and fabrication equipment; they are already coordinating, and no longer find the road-map process all that useful. "The chip manufacturer's buy-in is definitely less than before," says Chen.

Take the SRC, which functions as the US industry's research agency: it was a long-time supporter of the road map, says SRC vice-president Steven Hillenius. "But about three years ago, the SRC contributions went away because the member companies didn't see the value in it." The SRC, along with the SIA, wants to push a more long-term, basic research agenda and secure federal funding for it — possibly through the White House's National Strategic Computing Initiative, launched in July last year.

That agenda, laid out in a report<sup>5</sup> last September, sketches out the research challenges ahead. Energy efficiency is an urgent priority — especially for the embedded smart sensors that comprise the 'Internet of things,' which will need new technology to survive without batteries, using energy scavenged from ambient heat and vibration. Connectivity is equally key: billions of free-roaming devices trying to communicate with one another and the cloud will need huge amounts of bandwidth, which they can get if researchers can tap the once-unreachable terahertz band lying deep in the infrared spectrum. And security is crucial — the report calls for research into new ways to build in safeguards against cyberattack and data theft.

These priorities and others will give researchers plenty to work on in coming years. At least some industry insiders, including Shekhar Borkar, head of Intel's advanced microprocessor research, are optimists. Yes, he says, Moore's law is coming to an end in a literal sense, because the exponential growth in transistor count cannot continue. But from the consumer perspective, "Moore's law simply states that user value doubles every two years". And in that form, the law will continue as long as the industry can keep stuffing its devices with new functionality.

The ideas are out there, says Borkar. "Our job is to engineer them." ■

**M. Mitchell Waldrop** is a features editor for Nature.

1. Moore, G. E. *Electronics* **38**, 114–117 (1965).
2. Moore, G. E. *IEDM Tech. Digest* 11–13 (1975).
3. Sabry Aly, M. M. et al. *Computer* **48**(12), 24–33 (2015).
4. Nikolic, B. *41st Eur. Solid-State Circuits Conf.* (2015); available at <http://go.nature.com/wwwljik7>
5. *Rebooting the IT Revolution: A Call to Action* (SIA/SRC, 2015); available at <http://go.nature.com/urvkhw>



Scientists are becoming increasingly frustrated by how long it seems to take to publish papers. But is it really getting worse?

# ***THE WAITING GAME***

BY KENDALL POWELL



**W**hen Danielle Fraser first submitted her paper for publication, she had little idea of the painful saga that lay ahead.

She had spent some 18 months studying thousands of fossil species spread across North America from the past 36 million years, and now she had an intriguing result: animal populations were spread widest across latitudes in warm, wet climates. Her work, crucial to earning her PhD at Carleton University in Ottawa, Canada, might be used to make predictions about the response of mammals to climate change — a key question in ecology today. So, with her PhD adviser's encouragement, she sent it to *Science* in October 2012.

Ten days later, the paper was rejected with a form letter. She sent it to another prestigious journal, the *Proceedings of the National Academy of Sciences*. Rejected. Next, she tried *Ecology Letters*. Bounced. "At this point, I definitely was frustrated. I hadn't even been reviewed and I would've loved to know how to improve the paper," recalls Fraser. "I thought, 'Let's just get it out and go to a journal that will assess the paper.'"

In May 2013, she submitted the paper to *Proceedings of the Royal Society B*, considered a high-impact journal in her field. The journal sent it out for review — seven months after her initial submission to *Science*. "Finally!" Fraser thought. What she didn't know was that she had taken only the first steps down the long, bumpy road to publication: it would take another three submissions, two rejections, two rounds of major revisions and numerous drafts before the paper would finally appear. By that point, she could hardly bear to look at it.

Fraser's frustration is widely shared: researchers are increasingly questioning the time it takes to publish their work. Many say that they feel trapped in a cycle of submission, rejection, review, re-review and re-re-review that seems to eat up months of their lives, interfere with job, grant and tenure applications and slow down the dissemination of results. In 2012, Leslie Vosshall, a neuroscientist at the Rockefeller University in New York City, wrote a commentary that lamented the "glacial pace" of scientific publishing<sup>1</sup>. "In the past three years, if anything, it's gotten substantially worse," she says now. "It takes forever to get the work out, regardless of the journal. It just takes far too long."

But is the publication process actually becoming longer — and, if so, then why? To find out, *Nature* examined some recent analyses on time to publication — many of them performed by researchers waiting for their own work to see the light of day — and spoke to scientists and editors about their experiences.

The results contain some surprises. Daniel Himmelstein, a computational-biology graduate student at the University of California, San Francisco, analysed all the papers

indexed in the PubMed database that had listed submission and acceptance dates (see [go.nature.com/zjdqhn](http://go.nature.com/zjdqhn)). His study, done for *Nature*, found no evidence for lengthening delays<sup>2</sup>: the median review time — the time between submission and acceptance of a paper — has hovered at around 100 days for more than 30 years (see 'Paper wait'). But the analysis comes with major caveats. Not all journals — including some high-profile ones — deposit such time-stamp data in PubMed, and some journals show when a paper was resubmitted, rather

**"It takes forever to get the work out, regardless of the journal."**

than submitted for the first time. "Resetting the clock is an especially pernicious issue," Himmelstein says, and it means that the analysis might be underestimating publication delays.

Some data suggest that wait times have increased within certain subsets of journals, such as popular open-access ones and some of the most sought-after titles. At *Nature*, the median review time has grown from 85 days to just above 150 days over the past decade, according to Himmelstein's analysis, and at *PLoS ONE* it has risen from 37 to 125 days over roughly the same period.

Many scientists find this odd, because they expect advances in digital publishing and the proliferation of journals to have sped things up. They say that journals are taking too long to review papers and that reviewers are requesting more data, revisions and new experiments than they used to. "We are demanding more and more unreasonable things from each other," says Vosshall. Journal editors counter that science itself has become more data-rich, that they work to uphold high editorial and peer-review standards and that some are dealing with increasing numbers of papers. They also say that they are taking steps to expedite the process.

Publication practices and waiting times also vary widely by discipline — with social sciences being notoriously slow. In physics, the pressure to publish fast is reduced because of the common practice of publishing preprints — early versions of a paper before peer review — on the arXiv server. Some of the loudest complaints about publication delays come from those in biological fields, in which competition is fierce and publishing in prestigious journals can be required for career advancement. This month, a group of more than 70 scientists, funders, journal editors and publishers are meeting at the Howard Hughes Medical Institute campus in Chevy Chase, Maryland, to discuss whether biologists should adopt the

preprint model to accelerate publishing. "We need a fundamental rethinking of how we do this," Vosshall says.

## THE PITCH

In March 2012, Stephen Royle, a cell biologist at the University of Warwick, UK, started on a publication mission of his own. His latest work answered a controversial question about how cells sense that chromosomes are lined up before dividing, so he first sent it to *Nature Cell Biology* (NCB), because it is a top journal in his field and an editor there had suggested he send it after hearing Royle give a talk. It was rejected without review. Next, he sent it to *Developmental Cell*. Rejected. His next stop, the *Journal of Cell Biology*, sent the paper out for review. It came back with a long list of necessary revisions — and a rejection.

Royle and his lab spent almost six months doing the suggested experiments and revising the paper. Then he submitted the updated manuscript to *Current Biology*. Rejected. *EMBO Journal*. Reviewed and rejected.

Finally, in December 2012, he submitted it to the *Journal of Cell Science* (JCS), where it was reviewed. One reviewer mentioned that they had already assessed it at another journal and thought that it should have been published then. They wrote that the work was "beautifully conducted, well controlled, and conservatively interpreted". A second reviewer said that it should not be published. The editor at JCS decided to accept it. The time between first submission to *Nature Cell Biology* and acceptance at JCS was 317 days. It appeared online another 53 days later<sup>3</sup>. The work went on to win the JCS prize as the journal's top paper for 2013.

Despite the accolade, Royle says that the multiple rejections were demoralizing for his student, who had done the experiments and needed the paper to graduate. He also thinks that the paper deserved the greater exposure that comes from publication in a more prestigious journal. "Unfortunately, the climate at the moment is that if papers aren't in those very top journals, they get overlooked easily," he says. And Royle, who has done several publication-time analyses and blogged about what he found, has shown that this experience is not unusual. When he looked at the 28 papers that his lab had published in the previous 12 years, the average time to gestate from first submission to publication was the same as a human baby — about 9 months (see [go.nature.com/79h2n3](http://go.nature.com/79h2n3)).

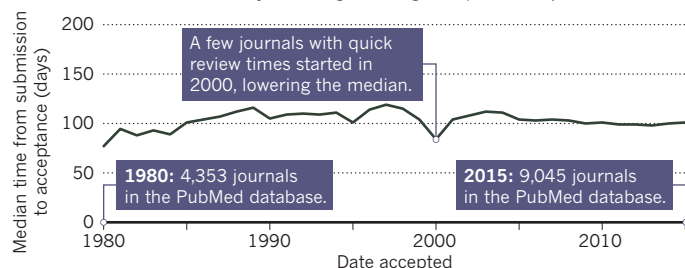
But how much of these delays were his own doing? To publish the chromosome paper, Royle indulged in the all-too-familiar practice of journal shopping: submitting first to the most prestigious journals in his field (often those with the highest impact factor) and then working his way down the hierarchy. (*Nature Cell Biology*'s current impact factor is 19; JCS's is 5.) Journal impact factor or reputation are widely used by scientists and grant-review and hiring committees as a proxy for the quality of the paper. On

# PAPER WAIT

Some scientists complain that publishing papers takes too long, but data show a complex picture.

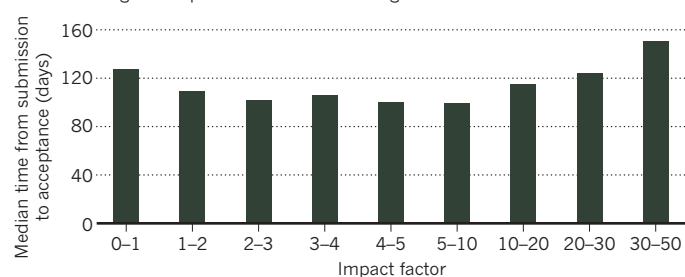
## REVIEW TIME

An analysis of all papers in PubMed up to 2015 with listed submission and acceptance dates suggests that the median time from submission to acceptance has hovered at around 100 days, although it has gone up at some journals.



## WAIT TIMES BY IMPACT FACTOR

An analysis of PubMed papers published in 2013 suggests that journals with the lowest and highest impact factors have the longest review times.



the flip side, critics say that editors seek out the splashiest papers to boost their publication's impact factor, something that encourages journal shopping, increases rejection rates and adds to the wait time. Journal editors reject this; Ritu Dhand, Nature editorial director in London, says that Nature's policy of selecting original, important work "may lead to citation impact and media coverage, but Nature editors aren't driven by those considerations".

How much time does journal shopping add? In the analysis of his group's research papers, Royle found that more than half were shopped around, and that this consumed anywhere from a few days to more than eight months. He went on to analyse all the papers published in 2013 that are indexed in PubMed, and examined whether higher impact factor correlated with longer median publication times. He found an inverted bell-shaped curve — the journals with the lowest and highest impact factors had longer review times than did those in the middle. For the vast majority of those in the middle, review times stood at around 100 days — matching Himmelstein's analysis. Those with the very highest impact factors (30–50) had a review time of 150 days, supporting the idea that pitching a paper to a series of top journals could result in significant delays in publication.

Many scientists, editors and publishers have long acknowledged that journal name is a flawed measure of the quality and value of a piece of research — but the problem shows no signs of going away. "Where your paper is published doesn't say anything about you, your paper's impact or whether it's right or wrong," says Maria Leptin, director of EMBO, an

organization of Europe's leading life scientists and publisher of the *EMBO Journal*. "Nobody has the courage to say, we, as a funding organization, or we, as a tenure committee, are not going to look at where you publish as opposed to what you publish."

And the obsession with prestigious journals is just one source of delay — as Fraser, who was battling to publish her paper on ancient animal populations — was about to found out.

## PEER REVIEW

By October 2013, a full year had passed since Fraser had first submitted her paper to a journal, and she had pretty much stopped caring about impact factor. By this point, the paper had spent two months in review at *Proceedings of the Royal Society B*, before coming back with mixed reports — and a rejection. So Fraser decided to try *PLoS ONE*, a journal that says it will publish any rigorous science, regardless of its significance, scope or anticipated citations. It has an impact factor of 3, and a reputation for rapid publication.

*PLoS ONE* sent the paper out to a single reviewer. Two months later, Fraser got a decision letter that essentially stated that the paper was rejected but might be eligible for re-review if the suggested revisions were made. She made the revisions, adding citations and a small amount of reanalysis. In March 2014, she resubmitted the manuscript, which *PLoS ONE* sent out to a different reviewer. Another two months passed before she received the new review: major revisions, please.

"I'm just happy they didn't tell me to go away," recalls Fraser. "I do have e-mails from the time

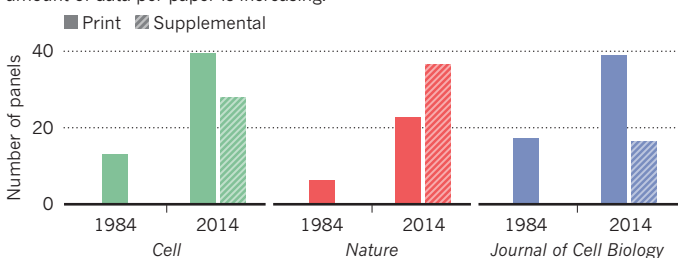
## PRODUCTION TIME

The same analysis of PubMed papers suggests that the time between acceptance and publication has dropped, probably because technology has improved.



## DATA PER PAPER

An analysis of biology papers published in 3 journals showed that the number of panels in experimental figures jumped between 1984 and 2014, a hint that the amount of data per paper is increasing.



that say, '1-millionth draft!' She persevered, making more revisions to meet the reviewer's demands, and in June 2014 submitted the paper to *PLoS ONE* for a third time. Success! The paper<sup>4</sup> was published online 23 months after she had first sent it to *Science*. The long peer-review and revision process did improve the paper, Fraser says now. "It was really much better." But did the main conclusion of the paper change? "Not really."

Last year, Chris Hartgerink, a behavioural sciences graduate student at Tilburg University in the Netherlands, ran an analysis of the Public Library of Science (PLOS) family of journals since the first one launched in 2003. (He chose the journals largely because they make the data easily accessible, and because he was waiting for a paper to be published in *PLoS ONE*.) He found that the mean review time had roughly doubled in the past decade, from 50–130 days to 150–250 days, depending on the journal (see [go.nature.com/s3voeq](http://go.nature.com/s3voeq)). And when Royle looked at eight journals that had published cell-biology papers over the past decade, he found that publication times had lengthened at seven of them, mostly because review times had stretched out.

One contention is that peer reviewers now ask for more. When Ron Vale, a cell biologist at the University of California, San Francisco, analysed biology papers that had been published in *Cell*, *Nature* and the *Journal of Cell Biology*, in the first six months of 1984 and compared that with the same period in 2014, he found that both the average number of authors and the number of panels in experimental figures rose by 2–4 fold<sup>5</sup>. This showed, he argued, that the amount of data required for a publication

SOURCES: REVIEW TIME: DANIEL HIMMELSTEIN; WAIT TIMES: STEPHEN ROYLE; PRODUCTION TIME: DANIEL HIMMELSTEIN; DATA: REF: 5



had gone up, and Vale suspects that much of the added data come from authors trying to meet reviewers' demands. Scientists grumble about overzealous critics who always seem to want more, or different, experiments to nail a point. "It's very rare for the revisions to fundamentally change a paper — the headline doesn't change," Royle says. His analysis of his group's publication times showed that almost 4 months of the average 9-month gestation was spent revising papers for resubmission.

Many scientists also blame journal editors, who, they say, can be reluctant to provide clear guidance and decisions to authors when reviews are mixed — unnecessarily stringing out the review and revision process. Journal heads disagree, and say that their editors are accomplished at handling mixed reviews. *Cell* editor-in-chief Emilie Marcus in Cambridge, Massachusetts, says that editors at her journal take responsibility for publication decisions and help authors to map out a plan for revisions.

Technological advances mean that research now involves handling more and more data, editors say, and there is greater emphasis on making that information available to the community. Marcus says that her journal is working to cut review times by, for example, increasing the number of papers that go through only one round of revision — 14% of their papers did so in 2015. In 2009, *Cell* also restricted the amount of supplemental material that could accompany papers as a way to keep requests for "additional, unrelated experiments" at bay.

PLOS executive editor Veronique Kiermer, based in San Francisco, declined to discuss the specifics of Fraser's paper, but she called its total review time of nine months an "outlier" and said that it was "not ideal to have research being evaluated by a single person". She acknowledges that *PLoS ONE*'s publication time has risen; one factor is that the volume of papers has, too — from 200 in 2006 to 30,000 per year now — and it takes time to find and assign appropriate editors and reviewers. (PLOS used 76,000 reviewers in 2015.) Another, says Kiermer, is that the number of essential checkpoints — including competing-interest disclosures, animal-welfare reports and screens for plagiarism — have increased in the past decade. "We'll do everything we can both in terms of technology and looking at workflows to bring these times to publication down," she says.

Dhand says that at *Nature*, too, editors find it harder to find reviewers than in the past, "presumably because there are so many more papers that need reviewing". Himmelstein found that the number of papers in PubMed more than doubled between 2000 and 2015, reaching nearly 1 million articles.

## TECHNOLOGY ADVANCES

Digital publishing may have had benefits in shortening 'production' time — the time from acceptance to publication — rather than time in review. In Himmelstein's analysis, time spent

in production has halved since the early 2000s, falling to a stable median of 25 days.

Several new journals and online publishing platforms have promised to speed up the process even more. PeerJ, a family of journals that launched in 2013, is one of several that now encourage open peer review, in which reviewers' names and comments are posted alongside articles. The hope is that the transparency will prevent unnecessary delays or burdensome revision requests from reviewers.

The biomedical and life-sciences journal *eLife* launched in 2012 with a pledge to make

**"It's very rare for the revisions to fundamentally change a paper."**

initial editorial decisions within a few days and to review papers quickly. Reviewers get strict instructions not to suggest the 'perfect experiment', and they can ask for extra analysis only if it can be completed within 2 months. Otherwise, the paper is rejected. Randy Schekman, a cell biologist at the University of California, Berkeley, and editor-in-chief of *eLife*, says that these policies mean that more than two-thirds of the journal's accepted papers undergo just one round of review.

In a 2015 analysis, Himmelstein created a ranking by the median review time for all 3,482 journals that had papers with time stamps in the PubMed database from January 2014 to June 2015 (see [go.nature.com/sscrr6](http://go.nature.com/sscrr6)). *PeerJ* had a relatively fast time: 74 days after submission. At *eLife*, it took 108 days, and *PLoS ONE* took 117. By comparison, *Cell*'s review time was 127 days; *Nature*'s was 173 days; *PLoS Medicine* took 177 days; and *Developmental Cell* was among the slowest of the popular biomedical journals, at 194 days. Marcus notes that comparison between journals is difficult because the publications define received, revised and accepted days differently, and that *Developmental Cell* places a high priority on timely review.

## PREPRINTS RECONSIDERED

One way for biologists to accelerate publication is by embracing preprints. These allow work to quickly receive credit and critique, says Bruno Eckhardt, associate editor of *Physical Review E* and a theoretical physicist at the University of Marburg in Germany. "It is almost like going on Facebook — it means you are ready to go public," he says. A preprint submitted to bioRxiv — a server run by Cold Spring Harbor Laboratory in New York — is published online within 24 hours and given a digital object identifier (DOI); subsequent revisions are time-stamped

and anyone can read and comment on the paper. "The minute a research story gets into the public domain, it benefits from the collective power of different brains looking at a problem," says Vale. What's more, proponents say, preprint publishing can simply be added onto the conventional publication process. *F1000Research*, which launched in 2012, does this by publishing papers first, then inviting open peer review and revision.

Some scientists are going a step further, and using platforms such as GitHub, Zenodo and figshare to publish each hypothesis, data collection or figure as they go along. Each file can be given a DOI, so that it is citable and trackable. Himmelstein, who already publishes his papers as preprints, has been using the Thinklab platform to progressively write up and publish the results of a new project since January 2015. "I push 'publish' and it gets a DOI with no delay," he says. "Am I really gaining that much by publishing [in a conventional journal]? Or is it better to do what is fastest and most efficient to get your research out there?"

But preprints and real-time digital publishing platforms are no panacea. Vossall says that many biologists are "terrified" of preprints because they fear getting scooped by competitors or losing credit and intellectual-property rights for their ideas. And even after preprint publishing, scientists can still find themselves sloggling through peer review and chasing high-impact journals for a final publication to adorn their CV. Vossall says that the scientific community relies on conventional journals to serve as a 'prestige filter' so that important papers are brought to the attention of the right readers. Without them, "How do we find the good stuff?" she asks.

For Fraser, her *PLoS ONE* publication proved a success. When the paper was finally published after its almost-two-year wait, she got positive responses, she says. It has been viewed nearly 2,000 times, had 51 shares on Facebook and Twitter and got 280 downloads. The publication also helped her to secure her current position — as a postdoctoral fellow at the Smithsonian Institution Museum of Natural History in Washington DC. "I pretty much got the top postdoc that I could have gotten."

Still, the whole process is not something she wants to endure again — so these days, she tends to send her papers to mid-range journals that are likely to publish her work right away. "If my ultimate goal is to get a faculty job, I can't afford to wait two years on a single paper," she says. ■

**Kendall Powell** is a writer based in Lafayette, Colorado.

1. Vossall, L. B. *FASEB J.* **26**, 3589–3593 (2012).
2. Himmelstein, D. S. & Powell, K. *Zenodo Repository* <http://doi.org/bb95> (2016).
3. Cheeseman, L. P. et al. *J. Cell Sci.* **126**, 2102–2113 (2013).
4. Fraser, D. et al. *PLoS ONE* **9**, e106499 (2014).
5. Vale, R. D. *Proc. Natl Acad. Sci. USA* **112**, 13439–13446 (2015).

# COMMENT

**LAW** Shareholders who don't drive emissions cuts may find themselves in court **p.156**

**DRUG DISCOVERY** How an immunologist became a philanthropist **p.157**

**COSMOLOGY** A personal tour of the 'sponginess' of the Universe **p.158**



**CLIMATE** Metrics needed to track adaptation as well as mitigation **p.160**

ROGELIO V. SOLIS/AP



Carbon-capture plants provide one way to reduce the amount of atmospheric carbon dioxide.

## Scrutinize CO<sub>2</sub> removal methods

The viability and environmental risks of removing carbon dioxide from the air must be assessed if we are to achieve the Paris goals, writes **Phil Williamson**.

In Paris last December, the 196 parties to the United Nations Framework Convention on Climate Change (UNFCCC) agreed to balance the human-driven greenhouse-gas budget some time between 2050 and 2100. This commitment is intended to limit the increase in global average temperature above pre-industrial levels to “well below 2°C” — and preferably to 1.5°C.

A balanced greenhouse-gas budget either

requires that industry and agriculture produce zero emissions or necessitates the active removal of greenhouse gases from the atmosphere (in addition to deep and rapid emissions cuts). In most modelled scenarios that limit warming to 2°C, several gigatonnes of carbon dioxide have to be extracted and safely stored each year<sup>1</sup>. For more ambitious targets, tens of gigatonnes per year must be removed<sup>2</sup>.

Many CO<sub>2</sub>-removal techniques have been proposed. Whether any of them could work at the scale needed to deliver the goal of the Paris agreement depends on three things: feasibility, cost and acceptability. A crucial component of all of these approaches is the non-climatic impacts that large-scale CO<sub>2</sub>-removal could have on ecosystems and biodiversity.

Until now, the UNFCCC's scientific advisory body, the Intergovernmental Panel on Climate Change (IPCC), has paid relatively little attention to such impacts. It has fallen to other groups to review insights and gaps in our understanding of the influence of CO<sub>2</sub>-removal techniques on ecology<sup>3–5</sup>; to make broad assessments of climate-engineering schemes<sup>6</sup>; and to carry out comparative modelling studies<sup>7</sup>.

It is time for the IPCC, governments and other research-funding agencies to invest in new, internationally coordinated studies to investigate the viability and relative safety of large-scale CO<sub>2</sub> removal.

### FRONT-RUNNERS

Since its establishment in 1988, the IPCC has predominantly involved physical scientists and modellers, rather than ecologists. This, combined with the only relatively recent evidence that emissions reduction alone is unlikely to avert dangerous climate change, could account for why the IPCC's roughly 5,000-page Fifth Assessment Report, released in 2013 and 2014, leaves out one crucial consideration: the environmental impacts of large-scale CO<sub>2</sub> removal.

This omission is striking because the set of IPCC emissions scenarios that are likely to limit the increase in global surface temperature to 2°C by 2100 (the aim of the RCP2.6 ‘representative concentration pathway’, the IPCC climate-change-response scenario that achieves the lowest emissions) mostly relies on large-scale CO<sub>2</sub> removal.

These scenarios assume that two techniques could be developed to balance the carbon budget later this century: bioenergy with carbon capture and storage (BECCS), and afforestation. BECCS involves growing bioenergy crops, from grasses to trees; burning them in power stations; stripping the CO<sub>2</sub> from the resulting waste gases; and compressing it into a liquid for underground storage. Afforestation — planting ▶



## TAKE YOUR PICK

A plethora of schemes have been proposed to extract carbon dioxide from the atmosphere. Here are nine, some more speculative than others.

### TECHNIQUE

### HOW IT WORKS

#### Bioenergy with carbon capture and storage (BECCS)



Crops grown for the purpose are burnt in power stations (providing energy), and the resulting CO<sub>2</sub> is captured for secure long-term storage.

#### Afforestation and reforestation



Large-scale tree plantations increase natural storage of carbon in biomass and forest soil.

#### 'Blue carbon' habitat restoration



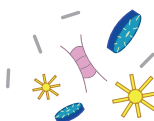
The recovery of degraded or over-exploited coastal ecosystems that have a high potential for carbon storage, such as saltmarshes and mangroves.

#### Biochar



Carbon from partly burnt biomass is added to soil, with potential for agricultural benefits.

#### Enhanced ocean productivity



Marine photosynthesis and CO<sub>2</sub> drawdown from the atmosphere is increased, either by adding nutrients to promote phytoplankton growth in the open ocean or through seaweed cultivation in shallow seas.

#### Enhanced weathering (using silicate rock)



Crushed olivine or other silicate rocks are added to soil surfaces or the ocean for chemical absorption of CO<sub>2</sub>. (Could help to reduce ocean acidification.)

#### Direct air capture (DAC)



Chemicals (or possibly low temperatures) are used to extract CO<sub>2</sub> from ambient air. Safe CO<sub>2</sub> transport and storage are subsequently required.

#### Cloud treatment to increase alkalinity



Alkaline rain resulting from cloud treatments reacts with, and removes, atmospheric CO<sub>2</sub>.

#### Building with biomass



A massive increase in the use of biomass (straw and timber) as a building material removes carbon for decades or centuries.

► trees — also relies on photosynthesis to initially remove CO<sub>2</sub> from the atmosphere. Storage is achieved naturally, in timber and soil.

Limiting the global temperature rise to 2°C, with any confidence, would require the removal of some 600 gigatonnes of CO<sub>2</sub> over this century (the median estimate of what is needed)<sup>8</sup>. Using BECCS, this would probably require crops to be planted solely for the purpose of CO<sub>2</sub> removal<sup>9</sup> on between 430 million and 580 million hectares of land — around one-third of the current total arable land on the planet, or about half the land area of the United States.

Unless there are remarkable increases in agricultural productivity, greatly exceeding the needs of a growing global population, the land requirements to make BECCS work would vastly accelerate the loss of primary forest and natural grassland. Thus, such dependence on BECCS could cause a loss of terrestrial species at the end of the century perhaps worse than the losses resulting from a temperature increase of about 2.8°C above pre-industrial levels<sup>10</sup>.

A more fundamental concern is whether BECCS would be as effective as it is widely assumed to be at stripping CO<sub>2</sub> from the atmosphere. Planting at such scale could involve more release than uptake of greenhouse gases, at least initially, as a result of land clearance, soil disturbance and increased use of fertilizer. When such effects are taken into account, the maximum amount of CO<sub>2</sub> that can be removed by BECCS (under the RCP2.6 scenario) is estimated to be 391 gigatonnes by 2100. This is about 34% less than the median amount assumed to be needed to keep the temperature rise below 2°C. If less optimistic but not unrealistic assumptions are made about where the land for bioenergy crops would come from, a net release of 135 gigatonnes of CO<sub>2</sub> could occur by 2100 (see 'Future unknown')<sup>8</sup>.

Incomplete understanding throws other assumptions of the BECCS-based scenarios into question<sup>9</sup>. For instance, little is known about the effect of future climatic conditions on the yields of bioenergy crops; what the water requirements of such crops may be in a warmer world; the implications for food security if bioenergy production directly competes with food production; and the feasibility (including commercial viability) of the associated carbon capture and storage infrastructure.

Less is expected of afforestation in terms of its ability to take CO<sub>2</sub> out of the atmosphere. Yet there is a near-universal assumption that increased forest cover is environmentally desirable. This is true in most cases of reforestation, particularly if a mixture of native trees is planted or replanted, rather than an exotic monoculture. But afforestation can also involve the loss of

natural ecosystems. And planting swathes of forest will cause complex changes in cloud cover, albedo (reflectance) and the soil–water balance (through changes to evaporation and plant transpiration), all of which affect Earth's surface temperature.

Counter-intuitively, afforestation at mid-latitudes and in northern, boreal forests may have a net warming effect, despite increasing the storage of carbon<sup>7</sup>. Also, as with bioenergy crops, it is difficult, if not impossible, to reliably quantify the effects of future climate change during 2050–2100. Increased fires, droughts, pests and disease could jeopardize the stability of carbon storage in newly planted forests.

### OTHER OPTIONS

There is no shortage of other ideas for CO<sub>2</sub> removal by biological, geochemical and chemical means (see 'Take your pick'). For all such schemes, modelling the theoretical potential of a proposed approach can give a completely different picture from that obtained when environmental impacts — not to mention practicalities, governance and acceptability — are considered.

The roughly 25 years of discussion, research and policymaking on ocean fertilization, another CO<sub>2</sub>-removal technique, is a case in point. Since the link was first made between natural changes in the input of dust to the ocean, ocean productivity and climatic conditions, there has been a dramatic scaling-down of expectations of how effective ocean fertilization might be as a way to avoid human-driven global warming<sup>11</sup>.

During the 1990s, researchers postulated that for every tonne of iron added to seawater, tens of thousands of tonnes of carbon (and hence CO<sub>2</sub>) could be fixed by the

resulting blooms of phytoplankton. This quantity has been whittled down over the years with the realization that most of the CO<sub>2</sub> absorbed

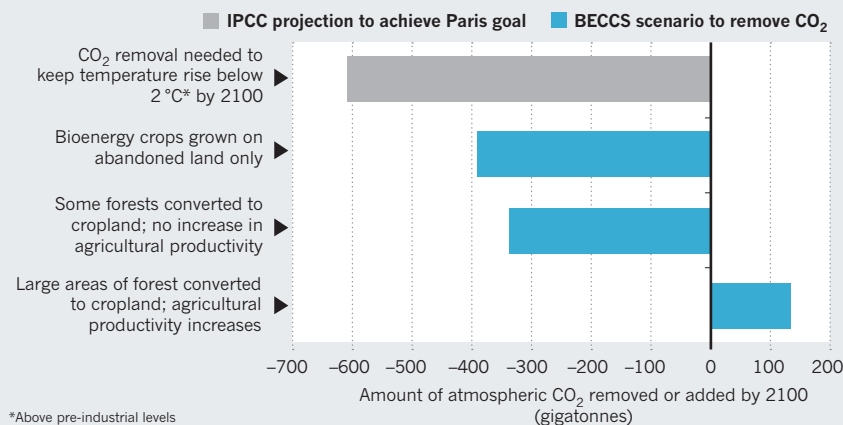
by such blooms — stimulated either by adding iron or other nutrients to seawater, or by enhancing upwelling through mechanical means — would be released back into the atmosphere when the phytoplankton decomposed. Moreover, a large-scale increase in plankton productivity in one region (across the Southern Ocean, say) could reduce the yields of fisheries elsewhere by depleting other nutrients, or increase the likelihood of mid-water deoxygenation. Such risks have resulted in the near-universal rejection of ocean fertilization as a climate intervention, through bodies such as the Convention on Biological Diversity (CBD)<sup>3</sup>.

More recently, other, potentially more controllable, ocean-based CO<sub>2</sub>-removal techniques have been suggested, such as the

**"Action should focus on urgent emissions reductions."**

## FUTURE UNKNOWN

Projections of how much carbon dioxide could be removed from the atmosphere using bioenergy with carbon capture and storage (BECCS) between 2020 and 2100 vary drastically depending on where the land for growing bioenergy crops comes from.



cultivation of seaweed to cover up to 9% of the global ocean<sup>12</sup>. The specific environmental implications of this method have yet to be assessed. Yet such an approach would clearly affect, and potentially displace, existing marine ecosystems that have high economic value. (Shallow and coastal waters currently provide around 90% of global fish catches.)

Back on land, other techniques include those to increase the amount of carbon sequestered in the soil, for example by ploughing in organic material such as straw, reducing ploughing (to limit soil disturbance) or adding biochar (a form of charcoal). Another idea is to enhance weathering, which involves the absorption of CO<sub>2</sub> from the atmosphere by certain silicate rocks. Existing insights from agriculture, geoscience and mineral extraction enable more informed assessments of the feasibility and acceptability<sup>3–6</sup> of these approaches. Yet it is crucial to know more about the permanence of carbon storage for biologically based methods, and the environmental impacts that might result if such approaches are used at vast scale<sup>4–6</sup>.

For example, the use of biochar raises land-use issues. In addition, millions of hectares of soil darkened by the application of biochar would decrease albedo, increasing heat absorption. The addition of pulverized rock to the soil surface, by contrast, would increase reflectivity. Yet to reduce the amount of CO<sub>2</sub> in the atmosphere by around 50 parts per million (a roughly 12% decrease from current levels), 1–5 kilograms per square metre of silicate rock would need to be applied each year to 2 billion to 6.9 billion hectares of land (15–45% of Earth's land surface area), mostly in the tropics<sup>13</sup>. The volume of rock mined and processed would exceed the amount of coal currently produced worldwide, with the total costs of implementation estimated to be between US\$60 trillion and \$600 trillion. And the

chemistry and biology of rivers and adjacent ocean areas would be radically altered.

The most environmentally benign option for large-scale CO<sub>2</sub> removal may be direct air capture (DAC). This can be done by passing air through anion-exchange resins that contain hydroxide or carbonate groups, which, when dry, absorb CO<sub>2</sub>, and release it when moist. The extracted CO<sub>2</sub> can then be compressed, stored in liquid form and deposited underground using carbon capture and storage technologies<sup>6</sup>.

The operational costs for DAC cover a similar range to those estimated for enhanced weathering. The extraction process would also need land and probably water, and, as for BECCS, there is a risk of CO<sub>2</sub> leaking out of geological reservoirs. Such risks can be minimized by storing the liquid CO<sub>2</sub> beneath the sea or by using geochemical transformation, which involves *in situ* reactions between CO<sub>2</sub> and certain rock types. In theory, cooling (rather than chemistry) to liquefy out the CO<sub>2</sub> could also be used to remove CO<sub>2</sub> from ambient air<sup>14</sup>. The technical feasibility, costs and potential environmental impacts of this approach — which could involve setting up plants in remote places such as Antarctica — have yet to be investigated.

## URGENT ACTION

As well as a major step up in research, urgent attention must be given to clarification at the UN level of what is considered geoengineering and what is climate mitigation. Once considered distinct approaches, the meaning of these terms has become fuzzier in recent years. CO<sub>2</sub> removal is frequently included in both categories, generating confusion and contradiction.

This is crucial to resolve because mitigation and geoengineering have very different psychological connotations. Mitigation is

universally considered to be a good thing that reduces risk or damage. Geoengineering frequently elicits suspicion, or is dismissed as a 'high-risk, high-tech' approach that may itself be harmful.

CO<sub>2</sub> removal was not specifically discussed in Paris. However, the large-scale extraction of CO<sub>2</sub> does seem to be a requirement to meet the goal of the Paris agreement. The CBD considers most, if not all, techniques for CO<sub>2</sub> removal to be climate geoengineering, which it has repeatedly rejected as a policy option for addressing climate change. With a few exceptions, the same 195 or so governments make up both the UNFCCC and the CBD.

One solution would be to abandon the term climate geoengineering and simply assess the various methods for mitigating climate change on a case-by-case basis.

The Paris agreement shows where we want to go — the brave new world of a balanced carbon budget — but not how to get there. For now, action should focus on urgent emissions reductions and not on an unproven 'emit now, remove later' strategy. But the unwelcome truth is that, unless a lot more effort is made to cut emissions, significant CO<sub>2</sub> removal will need to begin around 2020, with up to 20 gigatonnes of CO<sub>2</sub> extracted each year by 2100 to keep the global temperature increase 'well below 2°C'<sup>2</sup>.

Is that feasible? What environmental risks and constraints are involved? We need to know. ■

**Phil Williamson** is a science coordinator for the Natural Environment Research Council and an associate fellow in the School of Environmental Sciences at the University of East Anglia in Norwich, UK.  
e-mail: p.williamson@uea.ac.uk

1. Fuss, S. et al. *Nature Clim. Change* **4**, 850–853 (2014).
2. Rogelj, J. et al. *Nature Clim. Change* **5**, 519–528 (2015).
3. CBD Secretariat. *Update on Climate Geoengineering in Relation to the Convention on Biological Diversity* (CBD, 2015).
4. Smith, P. et al. *Nature Clim. Change* **6**, 42–50 (2016).
5. Smith, P. *Glob. Change Biol.* <http://dx.doi.org/10.1111/gcb.13178> (2016).
6. National Research Council. *Climate Intervention: Carbon Dioxide Removal and Reliable Sequestration* (National Academies Press, 2015).
7. Keller, D. P., Feng, E. Y. & Oschlies, A. *Nature Commun.* **5**, 3304 (2014).
8. Wilshire, A. & Davies-Barnard, T. *Planetary Limits to BECCS Negative Emissions* (AVOID2, 2015).
9. Gough, C. & Vaughan, N. *Synthesising Existing Knowledge on the Feasibility of BECCS* (AVOID2, 2015).
10. Newbold, T. et al. *Nature* **520**, 45–50 (2015).
11. Williamson, P. et al. *Process Safety & Environ. Protection* **90**, 475–488 (2012).
12. N'Yeurt, A. de R. et al. *Process Safety & Environ. Protection* **90**, 467–474 (2012).
13. Taylor, L. L. et al. *Nature Clim. Change* <http://dx.doi.org/10.1038/nclimate2882> (2015).
14. Agee, E., Orton, A. & Rogers, J. J. *Appl. Meteor. Clim.* **52**, 281–288 (2013).



# Shareholders must vote for climate-change mitigation

Investors who are standing idly by as emissions erode the value of their stock could find themselves in court, warn **Howard Covington** and colleagues.

The largest 500 companies listed on the world's stock markets account for about half of market value and 14% of global emissions<sup>1</sup>. Steel company ArcelorMittal, utilities firm RWE and oil giant ExxonMobil are among the top ten. All listed companies probably account for roughly one-quarter of global emissions. Many of these companies are multinationals, and are not party to the climate deal made in Paris last December, so they could resist or even challenge government plans to reduce emissions.

One way to encourage their cooperation would be for their shareholders to propose and vote for resolutions aimed at increasing or preserving stock value while reducing emissions. Why would shareholders do this? Senior lawyers have concluded that those who manage other people's money have a duty to control for 'material risks'<sup>2</sup>. In finance, that means risks that might trigger a 5% or more loss in investment value. Climate damage in the future is expected to be one such risk.

Therefore, we (an investment manager, an environmental lawyer and a climate economist) believe that clients and beneficiaries of investment firms might have a legal case to bring against their investors who stand idly by as emissions erode the value of their stock. We are researching and designing such actions.

## CASE LAW

Legal rulings about duties can have far-reaching consequences. In June 2015, a Dutch court found that the government has a duty of care to its citizens to minimize the risk of climate change, and ordered the Dutch government to go beyond its current plan to reduce the country's greenhouse-gas emissions. Pending an appeal hearing, the government is now working towards a 25% reduction by 2020.

In another incident last year, a case brought by ClientEarth (where one of us, J.T., works), the Supreme Court of the United Kingdom found that the British government has a duty to obey its own air-quality laws. Because this case involved the European Court, the rulings of which bind all European Union member states, ClientEarth is preparing a slew of parallel legal actions in other EU countries. Ten such cases were filed in Germany last November. One has already reached judgment, resulting in a court order to the city of Wiesbaden to take all necessary action to clean up

its air. The Chinese judiciary has sought views from ClientEarth on how legal action against emitters could help to address China's air and other pollution challenges.

Systemic climate risk for investors can be expressed in terms of the reduction in value, due to warming this century, of a diversified investment portfolio at a given probability. This is called a climate value at risk. If economic output is expected to be damaged by future warming<sup>3,4</sup>, then portfolio value may be lower than it otherwise would be. Economic damage could accelerate if extreme weather resulting from high levels of warming creates regional instabilities and causes large-scale migration by making populated areas uninhabitable<sup>5</sup>.

Assessment by the investment community of climate value at risk is still at an early stage<sup>6</sup>. Initial estimates point to a value at risk of 10% or so at probabilities of loss that are within the range typically used in financial management. The total value of the shares listed on the world's equity markets is around US\$70 trillion, suggesting that the potential loss from a plausible worst case is currently roughly \$7 trillion. This is a measure of the systemic climate risk to the world's equity markets.

## MANAGING RISK

This risk exceeds the legal test of materiality and should be too large to ignore. In practice, most investors neglect it entirely.

What should money managers be doing about systemic climate risk? Because it arises from damage to the economy as a whole, the risk cannot be reduced by hedging investments or by limiting exposure to particular assets. Investors should actively encourage the companies they own to reduce emissions, for example by urging profitable investment in energy efficiency or emissions-reducing changes to processes and by discouraging risky capital expenditure on fossil-fuel exploration and production.

Better still, managers could push companies to adopt the goals of the Paris agreement

**"Managers could push companies to adopt the goals of the Paris agreement into their constitutions."**

into their constitutions and to publish business plans for how they will deal with the transition to a zero-carbon economy.

A few forward-looking investors are already acting in this way and are beginning to have an effect. Last year, CCLA, a British investment firm, led the filing of shareholder resolutions requiring greater emissions disclosure from Shell and BP. This was followed by BHP Billiton, a mining company, responding to concerned investors by producing an analysis of how its business would operate in an emissions-constrained world. These are useful steps, but they are small.

To produce a wholesale change in attitude, a court ruling on the obligations of fiduciary investors to control systemic climate risk will probably be needed. Because of the uncertainties in estimating future climate damage, this will not be an easy case to bring. But we anticipate that such a case will ultimately succeed. We are seeking more and better quantitative assessments of systemic climate risk and, in particular, the risk to investment portfolios, to help to clarify the relevant obligations.

The Paris agreement was a diplomatic triumph. But it will remain largely voluntary until countries translate the promises into their own legal systems. Then it will take constant pressure to ensure implementation and prevent backsliding. Investors will play a major part, either voluntarily or because they will be forced by the courts to meet their legal obligations to manage climate risk. ■

**Howard Covington** is chair of the Alan Turing Institute in London. **James Thornton** is chief executive of ClientEarth, an environmental-law firm in London. **Cameron Hepburn** is professor of environmental economics at the University of Oxford, UK.  
e-mail: [hcovington@turing.ac.uk](mailto:hcovington@turing.ac.uk)

1. Moorhead, J. & Nixon, T. *Global 500 Greenhouse Gases Performance 2010–2013: 2014 Report on Trends* (Thomson Reuters, 2014).
2. Law Commission. *Fiduciary Duties of Investment Intermediaries* (UK Law Commission, 2014).
3. Burke, M., Hsiang, S. M. & Miguel, E. *et al. Nature* **527**, 235–239 (2015).
4. Dietz, S. & Stern, N. *Econ. J.* **125**, 574–602 (2015).
5. Pal, J. S. & Eltahir, E. A. B. *Nature Clim. Change* **26**, 197–200 (2016).
6. Economist Intelligence Unit. *The Cost of Inaction: Recognising the Value at Risk from Climate Change* (Economist Intelligence Unit, 2015).

**Love and Science:  
A Memoir**  
JAN VILCEK  
Seven Stories: 2016.

from the autoimmune drug Remicade (infliximab) to honour the contribution of immigrant scientists

and artists to US society.

Vilcek begins with how he discovered the drug. He cut his research teeth on interferons — proteins produced in response to infection — at the Bratislava Institute of Virology in the early 1960s. But it was in 1988 at New York University's School of Medicine that the main event occurred: the development of a monoclonal antibody against the inflammation-inducing molecule tumour necrosis factor (TNF). As Vilcek and others revealed, people with rheumatoid arthritis had high levels of TNF. Remicade was the first TNF-blocking agent approved by the US Food and Drug Administration, initially for treating Crohn's disease, in 1998. Today, it is used against multiple autoimmune conditions, including rheumatoid arthritis and psoriasis, and remains one of the world's five top-selling prescription drugs.

Vilcek hymns the importance to medicine of fundamental science, serendipity, collegiality and some risk-taking. He supports funding for translational medicine, but cautions that without continued support for basic research, there might be nothing to translate. However, the book's early detail about science, clinical trials and intellectual-property processes behind Remicade could be intimidating to those who picked up the autobiography for its personal tale. *Love and Science* becomes more engaging in its second section, when it steps back to the memories of a Jewish boy born in the year that Adolf Hitler came to power in Germany.

It is a tale of ingenuity and near misses. Vilcek's mother officially converted to Catholicism in 1939, his father in 1942, allowing both to be exempted from the Czechoslovakian government's punitive measures against Jews. They placed Jan in a Catholic orphanage at the age of eight, in a further attempt to keep him from being deported. Through several moves and periods of separation, the family survived the Second World War, and re-established a comfortable lifestyle in Bratislava in 1945. The democratic republic of Czechoslovakia had been reinstated after years of occupation and division during the war, but upheaval returned in 1948 when the Communist Party gained political power and quickly imposed a totalitarian system.

Vilcek decided early on to pursue research. He joined the Institute of Virology immediately after graduating from medical school in 1957, and soon published a single-author paper in *Nature* (J. Vilcek *Nature* **187**, 73–74; 1960). The young scientist doggedly established contacts with Western researchers, and these efforts paid off when he and Marica defected in 1964. Carrying no ►



Jan Vilcek at medical school in Bratislava in the 1950s.

#### DRUG DISCOVERY

# A life of tumult and triumph

**Marian Turner** reviews the memoir of émigré virologist and millionaire philanthropist Jan Vilcek.

**M**edical students in communist Czechoslovakia in the early 1950s were taught that Mendelian genetics was “bourgeois pseudoscience”. The theory of heredity that toed the party line came from Soviet agronomist Trofim Lysenko: Stalin championed his idea that acquired traits could be inherited. Jan Vilcek was one of several students of the time who saw through this nonsense, but kept quiet. The budding virologist already had two decades of experience of staying under the authoritarian radar — a frustration that would later prompt him

to break free in spectacular fashion.

In *Love and Science*, Vilcek traces that tumultuous, ultimately triumphant journey. He takes us from his birth in 1933, to Jewish parents in Bratislava (in what is now Slovakia), to his current position of president and co-founder, with his art-historian wife

Marica, of the Vilcek Foundation in New York City. This multi-million-dollar philanthropic enterprise uses Vilcek's royalties

► **NATURE.COM**

For more on science  
in culture see:  
[nature.com/  
booksandarts](http://nature.com/booksandarts)



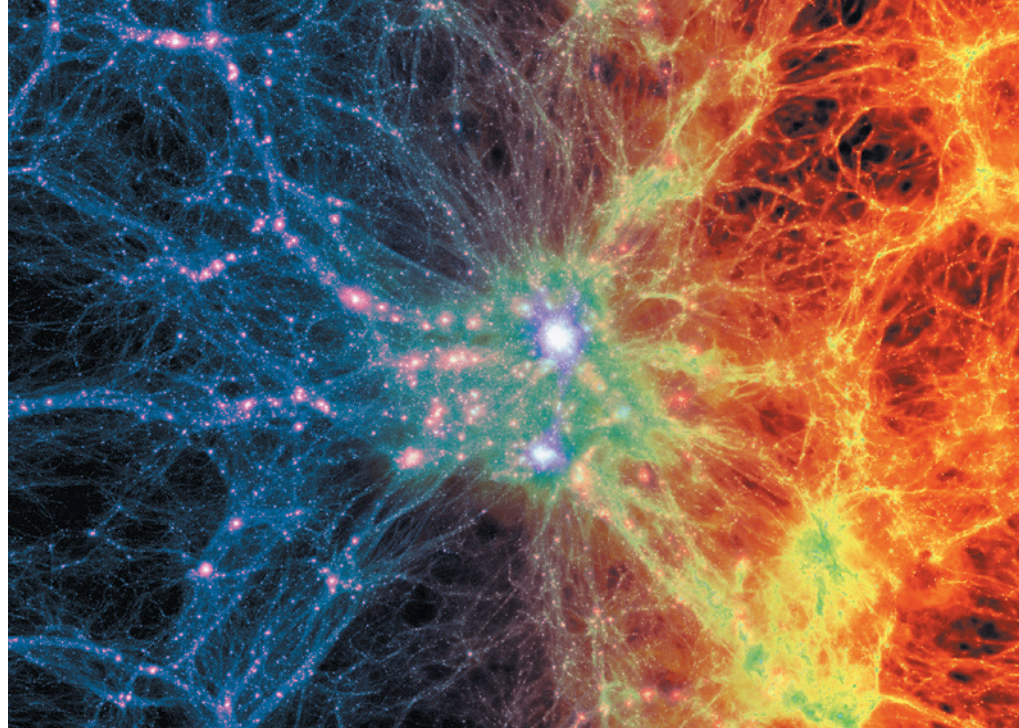
► official documents, they were waved across the border from Austria to West Germany — a scene rendered poignant in light of the current European refugee crisis. The ease with which Vilcek procured job offers from US colleagues is also hard to imagine now. He joined New York University as an assistant professor with no interview, a beneficiary of the US government's competitive research investments during the cold war.

The final section of the book I found the most original. Here, Vilcek reverts to the story of Remicade's success, and his unexpected wealth. His first quarterly royalty payment in 1999 was just less than his annual salary as a professor; by 2005, the portion of his future royalties pledged to New York University was projected to eventually reach more than US\$105 million. At first, Jan and Marica ate at restaurants and caught taxis more often; they helped family and friends, and replaced their second-hand furniture with European art-deco pieces. But they had no interest in luxury living, so as the royalties grew, they established what would become the Vilcek Foundation in 2000.

They began by supporting their workplaces, the New York University School of Medicine and the Metropolitan Museum of Art. They endowed professorships, lab space, scholarships and curatorships, earning a place among the top 15 US philanthropists of 2005. After the terrorist attacks of 11 September 2001, they began to see chinks in the US welcome for foreigners. Accordingly, in 2006 they launched annual Vilcek Prizes to recognize extraordinary achievements by immigrants to the United States in biomedical science and the arts and humanities. Aware that many of their prizewinners, from cancer biologist Joan Massagué to cellist Yo-Yo Ma, were already lauded, the Vilceks established annual prizes "for creative promise" in 2009 to honour up-and-coming talent. Prizewinners so far have included the Iranian-American scientist Pardis Sabeti, for her work on pathogen evolution (see N. L. Yozwiak *Nature* **518**, 477–479; 2015).

This story is told humbly, with honest insight into the deliberations between the Vilceks and generous credit to the many people who have advised them. The couple clearly delights in following the careers of the awardees. Jan Vilcek is now in his early eighties, and he mentions in *Love and Science* how he and Marica have made provision for the foundation to continue when they can no longer administer it. One senses that he is happy with his legacy, both scientific and benevolent. ■

**Marian Turner** is senior News & Views editor for Nature.



Density fluctuations in the Universe (shown in a computer simulation) have a sponge-like topology.

#### ASTRONOMY

## Topology quest

**Michael Blanton** enjoys a history of cosmology focused on large-scale structure in the 'spongy' Universe.

Over the past half-century, astrophysicists and cosmologists have revealed how structure formed in the Universe. This has vast importance: we exist only because tiny initial fluctuations in the density of matter grew through gravity to form galaxies, stars and planets. Key pieces of physics, astronomy and mathematics have only in recent decades been combined into a scientifically coherent and convincing origin story. In *The Cosmic Web*, astrophysicist Richard Gott, a compelling teller of cosmological stories, describes his part in this scientific quest.

He begins with a brisk primer on the early history of cosmology, starting with the discovery of the expansion of the Universe by US astronomer Edwin Hubble and others in the 1920s. The second half of the twentieth century, he shows, brought the detection by Arno Penzias and Robert Wilson of the cosmic microwave background radiation left over from the very early Universe (predicted by George Gamow and others), and the realization that fluctuations in this radiation would reveal the irregularities in density that seeded the growth of large-scale cosmic structure.

This is when Gott's work enters the picture. Beginning with his high-school obsession, topology — how geometric shapes can be connected — the book comes into focus. We start to follow Gott's fascinating path towards a better understanding of large-scale cosmic structure. Remarkably, his early discoveries

led to a paper in *American Mathematical Monthly* when Gott was just 20, on a new class of regular 'pseudopolyhedrons' with sponge-like topology (J. R. Gott *Am. Math. Mon.* **74**, 497–504; 1967). These polyhedrons are infinite, and their surfaces divide space into two identically shaped, interlocked regions. As in a sponge, each region (the 'sponge' and the 'holes' in it) is fully connected. If the Universe is separated into regions greater than and less than the mean density, it shows the same sponge-like topology, although with irregular patterns. Gott's enduring (and endearing) enthusiasm for this connection carries the reader through the more technical parts of the book.

*The Cosmic Web* is not a complete or representative intellectual history of twentieth- or twenty-first-century cosmology. Gott is rigorous in assigning credit for findings, but tells the history through the lens of his collaborations. This could leave a reader with the impression that a relatively small group of scientists is responsible for the discovery of dark matter, the development



**The Cosmic Web: Mysterious Architecture of the Universe**  
J. RICHARD GOTT  
Princeton University Press: 2016.

ILLUSTRATION

of structure-formation theory and progress on unravelling galaxy formation. Gott's major protagonists (including James Gunn, Jeremiah Ostriker and James Peebles) did make key discoveries, but in the context of a broader group than Gott describes. Gott also gives only a partial account of the creation and development of a key contemporary effort: work on large-scale structure and mapping of the cosmic web by the Sloan Digital Sky Survey (SDSS) — which uses the optical telescope at Apache Point Observatory, New Mexico — to create the most detailed 3D maps of the Universe yet, and continues to map further regions and earlier times.

Gott's is a personal view of cosmological developments, giving space and life to realities of research. This evokes the best of physicist and science historian Abraham Pais's more comprehensive books, such as *Inward Bound* (Clarendon/Oxford Univ. Press, 1988). Unafraid of spending a page explaining an idea that turns out to be wrong, Gott engagingly describes the blind alleys along the way to current views. He presents figures and equations stripped to their fundamental forms, but avoids the common temptation to water them down unrecognizably.

His work shines with this approach. As 3D maps of the Universe emerged from the 1980s onwards, revealed by surveys such as the SDSS and its forerunners, and as the theoretical picture of cosmology developed,

**"Gott's is a personal view of cosmological developments, giving space and life to realities of research."**

it became clear that the pattern of density fluctuations in the Universe is a cosmic, topologically sponge-like web. Gott's youthful work had primed him to grasp this, and gave him the

mathematical tools to test it. This aspect of cosmology is now an essential facet of our knowledge, and our fullest understanding of the topology of large-scale structure is derived from techniques that Gott developed.

Gott's journey shows how scientists can be so motivated by their earliest obsessions that they persist in pursuing them — and how unique obsessions can let them bring something new to the crowded table of ideas. For working scientists, this book is a reminder of what drives us, the value of chasing down questions that only we would ask, and how circuitous that chase can be. *The Cosmic Web* is not just a well-told story about the frontiers of cosmological knowledge. It is also an inspiration to explore them further. ■

**Michael Blanton** is an associate professor in the Center for Cosmology and Particle Physics at New York University, and is director of the Sloan Digital Sky Survey IV. e-mail: michael.blanton@gmail.com

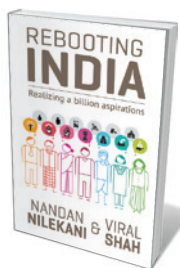
## Books in brief



### In a Different Key: The Story of Autism

John Donovan and Caren Zucker CROWN (2016)

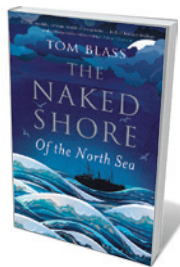
Hot on the heels of Steve Silberman's *Neurotribes* (Avery, 2015; see *Nature* **524**, 288–289; 2015) comes another monumental history of autism. Journalists John Donovan and Caren Zucker (whose television series *Echoes of Autism* broke ground in the 2000s) trace the halting progress in understanding the condition, weaving in stories of, among others, Donald Triplett, the first person diagnosed as autistic; US medic Leo Kanner; and UK psychiatrist Lorna Wing. The scoop is historian Herwig Czech's claim that key player Hans Asperger contributed to Nazi 'social cleansing' of children deemed mentally ill.



### Rebooting India: Realizing a Billion Aspirations

Nandan Nilekani and Viral Shah ALLEN LANE (2016)

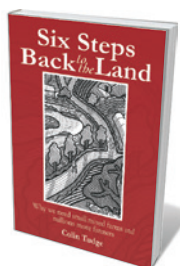
India's telecoms network is the world's second largest, yet the country's rural poor often effectively run marathons in accessing services. In this audacious technological manifesto — now in a UK edition — entrepreneur Nandan Nilekani and software specialist Viral Shah argue that India's vast challenges in banking, health care and other systems are solvable, if approached like government start-ups run by handpicked entrepreneurs. Nilekani and Shah cite the e-Aadhaar biometrics identification card, open to all Indians, as a first step on the road to digitally streamlined governance and society.



### The Naked Shore: Of the North Sea

Tom Blass BLOOMSBURY (2016)

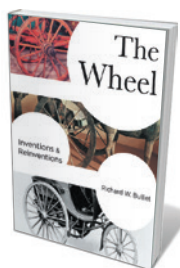
Compared to the charismatic Mediterranean, Europe's North Sea can seem a workaday tract of "mists, miasmas and surliness", notes journalist Tom Blass. But in trawling its depths for this vivid travelogue, Blass dredges up nuggets from an eventful cultural, military, industrial, economic and ancient history. Launching from the estuarine murk of the Thames, he takes in Dutch polders (tracts of land dubbed 'waking', 'sleeping' or 'dreaming', depending on distance from the sea), Frisian dialects, fishing fracas, puffins, porpoises and a future of warming waters and giant wind farms.



### Six Steps Back to the Land: Why We Need Small Mixed Farms and Millions More Farmers

Colin Tudge GREEN (2016)

The much-chewed-over conundrum of how to feed a projected global population of 10 billion demands an agrarian renaissance led by scientifically run, small, mixed farms, avers biologist Colin Tudge. Using principles such as agro-ecology and a focus on "lots of plants, not much meat, and maximum variety", Tudge's transformative prescription is fresh, pragmatic and packed with cutting-edge science. It is impressive, too, for its sound circular-economy thinking, not least an emphasis on an informed food culture.



### The Wheel: Inventions and Reinventions

Richard W. Bulliet COLUMBIA UNIVERSITY PRESS (2016)

As innovations go, the wheel might seem the ubiquitous driver of cultural evolution. But Richard Bulliet's technological history reveals complexities. The wheel was invented 3 times (with fixed and moving wheels on axles both emerging 5,000 years ago, and casters a mere 300); defined differently by the rail and automobile industries; and eschewed by civilizations such as the pre-Columbian Olmecs. Hitched in turn to rickshaws, ox carts, Mongolian mobile homes, barouches and bicycles, this is a deft narrative. **Barbara Kiser**



# Correspondence

## When brain bullets met crowdfunding

Stereotaxy — a surgical technique that uses a head clamp to pinpoint brain areas from coordinates of external landmarks and anatomical structures — was first applied in the nineteenth century, when it went largely unnoticed by the medical community. It was ‘reinvented’ 50 years later and is routinely used today, for example in deep-brain stimulation for diseases such as Parkinson’s.

Gaston Contremoulins (1869–1950) was the self-educated physicist who invented the ‘metroradiographic’ frame that marked the birth of stereotactic surgery (pictured). It was first used in 1897 to guide the removal of bullets from the brains of two young men who had survived after shooting themselves in the head.

This surgical triumph was crowdfunded following a public appeal by Contremoulins. It was reported in the French weekly newspaper *l’Illustration* and aroused great popular interest. In the scientific press, it attracted just a single report — in the proceedings of the French Academy of Sciences, *Compte-rendus Hebdomadaires de l’Académie des Sciences*.

**Caroline Apra, Pierre Bourdillon, Marc Lévêque**  
*Pitié-Salpêtrière Hospital, Paris, France.*  
[marc.leveque@aphp.fr](mailto:marc.leveque@aphp.fr)

## IPBES disciplinary gaps still gaping

The Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES) convenes this month to approve summaries for policymakers of the first assessments at its fourth plenary.

Despite early calls for IPBES to draw on a broader range of disciplines than did the Intergovernmental Panel on Climate Change (see E. Turnhout *et al.* *Nature* **488**, 454–455; 2012), the social sciences and



the humanities remain markedly under-represented. They make up less than 10% of the membership of IPBES expert groups, instead of the recommended 30%. These disciplines should play a bigger part in IPBES assessments and in implementing the first IPBES work programme for 2014–18.

The imbalance mirrors institutional and knowledge barriers between research disciplines. The IPBES Secretariat and its Multidisciplinary Expert Panel need to consult more experts from the social sciences and humanities for nominations for assessments. The panel should encourage these stakeholders to engage in scoping and reviewing activities and to register on IPBES networking sites.

One of the IPBES objectives is to include experts with “balance in the terrestrial and marine natural sciences, social and economic sciences, and arts and humanities”. The secretariat should review the disciplinary balance of all IPBES activities and products, and make the findings publicly available.

**Alice B. M. Vadrot** *University of Cambridge, UK.*

**Jens Jetzkowitz** *Philipps University Marburg, Germany.*

**Lindsay C. Stringer** *University of Leeds, UK.*  
[l.stringer@leeds.ac.uk](mailto:l.stringer@leeds.ac.uk)

## More to fisheries than catch limits

You suggest that the European Union is setting a worrying trend by ignoring scientific advice on overfishing, but that is only part of the story (*Nature* **528**, 435; 2015).

The advice you mention refers to catch levels that are in line with the objectives of the EU Common Fisheries Policy. Those are based on the fish-stock biomass that can deliver maximum sustainable yields. Scientific advice on the annual catch limits (known as total allowable catches, or TACs) is aimed at delivering maximum sustainable yield and is provided by the International Council for the Exploration of the Sea (ICES). However, the TACs are set by the EU Council of Ministers and, as you point out, can sometimes exceed the levels advised by ICES.

Nevertheless, to accuse the EU of “ignoring” scientific advice in such cases oversimplifies the decision-making process. In fact, the extent to which TACs have exceeded the levels recommended by ICES has decreased since 2001, as you note.

Other objectives in the fisheries policy relate to their economic and social benefits. They aim to ensure the availability of food supplies and to contribute to a fair living standard for those who are

dependent on fishing. Decision-making for fisheries management is therefore more complex than simply setting catch limits.

**John Casey, Jann T.**

**Martinsohn, Hendrik Dörner**  
*European Commission — JRC*  
*Institute for the Protection and Security of the Citizen, Ispra, Italy.*  
[john.casey@jrc.ec.europa.eu](mailto:john.casey@jrc.ec.europa.eu)

## Metrics needed to track adaptation

A lesser-known success of the Paris climate agreement in 2015 is the ‘global adaptation goal’, an ambitious plan for adapting to climate change that reaches beyond national boundaries. This is important because climate-change mitigation needs to take the world’s adaptation potential into account.

To sustain the long-term goal of keeping the global average temperature rise well below 2°C, we also need evidence that the world can adapt to the impacts of warming. The Paris agreement aims to build a collective understanding of adaptation through metrics and tools that capture each country’s efforts. Aggregating national contributions as a global trend will indicate whether humankind is on track to adapt.

The first step will be the agreement’s ratification in April by at least 55 parties to the United Nations Framework Convention on Climate Change (UNFCCC) — which together account for at least 55% of total greenhouse-gas emissions.

It will next be necessary to define metrics that both reflect national circumstances and allow aggregation. To this end, scientists tracking adaptation will need to work with experts at the UNFCCC and organizations such as the UN Environment Programme.

**Alexandre K. Magnan** *Institute for Sustainable Development and International Relations, Paris, France.*  
[alexandre.magnan@iddri.org](mailto:alexandre.magnan@iddri.org)

## SCHIZOPHRENIA

## From genetics to physiology at last

The identification of a set of genetic variations that are strongly associated with the risk of developing schizophrenia provides insights into the neurobiology of this destructive disease. [SEE ARTICLE P.177](#)

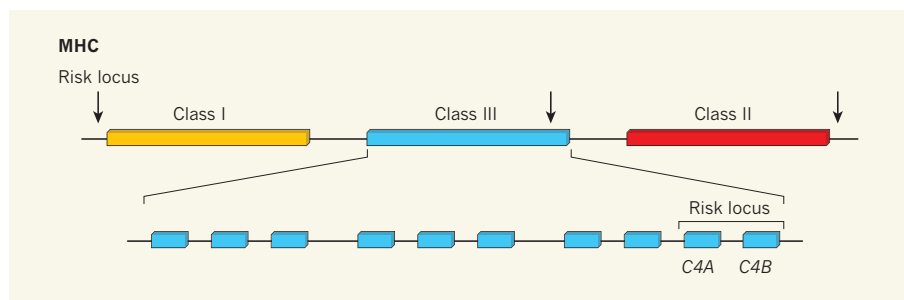
RYAN S. DHINDSA & DAVID B. GOLDSTEIN

Schizophrenia is a devastating and chronic neuropsychiatric disorder that affects nearly 1% of the world's population. It has long been hoped that the identification of genetic risk factors for schizophrenia would help to identify the physiological causes of the disease, but, despite decades of intensive research, the biological underpinnings of schizophrenia have remained elusive. On page 177 of the issue, Sekar *et al.*<sup>1</sup> present a remarkable genomic and neurobiological study that finally delivers on this long-standing hope.

There are two typical paths to drug discovery. The first is serendipity — sometimes researchers simply get lucky and find compounds that happen to work. Although such fortunate accidents have led to the development of some of medicine's most important drugs, this method of discovery cannot be systematized, and the resulting treatments often fall well short of a cure. Current treatments for schizophrenia fall uniformly into this category. The second, preferable, path is to understand the mechanisms that underpin a disease, and to design treatments that target those underlying causes.

One of the main problems in deciphering the molecular basis of schizophrenia is that there is a near complete absence of clearly associated biological changes. However, schizophrenia is a strongly genetic disorder, and for decades many have believed that understanding the genetics involved might provide a way to begin to dissect the biology of the disease. Until now, following this line of reasoning has been a largely frustrating experience. Before the advent of genome-wide association studies (GWAS), which link disease risk to specific regions of the genome, the field was plagued by a lack of statistical rigour and false starts, as researchers reported hundreds of putative associations between genetic variations and the risk of schizophrenia<sup>2</sup>. But GWAS have demonstrated that these early associations were largely incorrect<sup>3,4</sup>.

Careful GWAS have now finally identified real associations between genetic variants and schizophrenia. In particular, a 2014 study that used a huge sample size and strict statistical methods revealed that there are at least 108 genetic regions (loci) associated with a risk



**Figure 1 | A complex association.** Variations in numerous regions of the genome have been associated with an increased risk of schizophrenia, with many of the strongest associations occurring in the major histocompatibility complex (MHC). This region is challenging to study because it is 3.6 million bases long and complex, containing three classes of gene that have differing roles in immunity. Genetic loci that are associated with risk are found along the entire region (only three of the strongest associations are shown here, for simplicity). However, genetic variants in *C4A* and *C4B* are most strongly associated with increased risk. Sekar *et al.*<sup>1</sup> overcame the challenges of studying the MHC to pinpoint how variations in the *C4* genes that alter gene expression can increase risk.

of schizophrenia<sup>4</sup>. However, finding a locus is not the same as identifying a causal gene — for instance, loci are often located in sequences that don't encode genes, and many different variations can occur in one region, making it hard to pinpoint what exactly is driving the risk signal. Indeed, these 108 risk factors were not traced to specific genes or variant sequences. Without that information, a clear insight into the molecular aetiology of the disease remained lacking, but Sekar *et al.* have now taken that crucial next step.

The strongest risk association for schizophrenia is found in the major histocompatibility complex (MHC) — a region on chromosome 6 that contains genes involved in acquired immunity (Fig. 1). Within the MHC, the strongest risks are associated with loci near the gene *C4*, which encodes a complement factor (a part of the innate immune system). This association motivated the authors to ask whether variations in *C4* might be involved in schizophrenia risk.

In humans, *C4* exists as two distinct genes, *C4A* and *C4B*, and the number of copies of each gene varies from person to person. To complicate things further, there are long and short versions of *C4A* and *C4B*, which are determined by the presence or absence of a human endogenous retroviral (HERV) insertion in a non-coding region of the gene. Sekar *et al.* developed a way to accurately assess

the number of copies of each gene, and then used RNA-expression data to relate gene copy number and HERV status to *C4A* and *C4B* gene-expression levels.

With these data in hand, the authors present compelling evidence that they have identified the underlying cause of one of the main MHC risk associations. They have found that variation at *C4* is associated with schizophrenia risk independently of other genetic variations in the MHC region. Specifically, different levels of risk are associated with different genetic combinations of *C4* copy number and HERV status. The authors used RNA data to show that expression of *C4A* and *C4B* increases with copy number, and that the presence of the HERV insertion increases the ratio of *C4A* to *C4B* expression. Moreover, the higher the levels of *C4A* expression, the greater the risk of schizophrenia.

It is worth emphasizing that defining causal variants in the MHC regions is one of the most challenging problems in human statistical genetics. The MHC is 3.6 million bases long, and variants associated with schizophrenia risk have been found across the entire region. Indeed, the region is so challenging that geneticists often joke that they hope to find associations anywhere except in the MHC. Caution is imperative when trying to pinpoint causation in such a genetically complex region. However, Sekar and colleagues seem



to have successfully overcome these challenges.

The complement pathway has been implicated in synaptic pruning<sup>5</sup>, a developmental process in which the synaptic connections between neurons are continuously eliminated in the brain until early adulthood. Using a mouse model, Sekar *et al.* found that *C4* expression is upregulated during periods of synaptic pruning. By contrast, mice deficient in *C4* showed signs of decreased pruning. Thus, the authors postulate that increased *C4A* expression in individuals with schizophrenia results in increased synaptic pruning. Interestingly, studies of the brains of humans with schizophrenia have shown that affected individuals exhibit thinning and reduced synaptic structures in the cortical region of the brain compared with people without the disorder<sup>6–8</sup>. Hyperactive synaptic

pruning might explain these findings.

Unfortunately, because mice lack the two forms of *C4* found in humans, the question of why schizophrenia risk depends specifically on *C4A* expression levels remains open. Answering this question is now a priority, and can be expected to provide further mechanistic information.

Sekar and colleagues' study finally gives us the first real inroad into the molecular aetiology of schizophrenia, and perhaps a direction for the development of treatments. Although pruning undoubtedly represents a challenging therapeutic target, the authors' beautiful and comprehensive study gives much-needed inspiration for all those researchers who are trying to leverage genetics to advance our understanding of the biology of neuropsychiatric diseases. ■

**Ryan S. Dhindsa and David B. Goldstein**  
are at the Institute for Genomic Medicine,  
Columbia University Medical Center,  
New York, New York 10032, USA.  
e-mails: [rsd2135@cumc.columbia.edu](mailto:rsd2135@cumc.columbia.edu);  
[dg2875@cumc.columbia.edu](mailto:dg2875@cumc.columbia.edu)

1. Sekar, A. *et al.* *Nature* **530**, 177–183 (2016).
2. [www.schizophreniaforum.org/res/sczgene](http://www.schizophreniaforum.org/res/sczgene)
3. The International Schizophrenia Consortium. *Nature* **460**, 748–752 (2009).
4. Schizophrenia Working Group of the Psychiatric Genomics Consortium. *Nature* **511**, 421–427 (2014).
5. Stevens, B. *et al.* *Cell* **131**, 1164–1178 (2007).
6. Cannon, T. D. *et al.* *Biol. Psychiatry* **77**, 147–157 (2015).
7. Glausier, J. R. & Lewis, D. A. *Neuroscience* **251**, 90–107 (2013).
8. Glantz, L. A. & Lewis, D. A. *Arch. Gen. Psychiatry* **57**, 65–73 (2000).

This article was published online on 27 January 2016.

## CLIMATE SCIENCE

# A great Arctic ice shelf

**Newly mapped features on the floor of the Arctic Ocean suggest that the Arctic basin was once covered by a one-kilometre-thick, flowing ice shelf derived from large ice sheets in eastern Siberia, Arctic Canada and the Barents Sea.**

EUGENE DOMACK

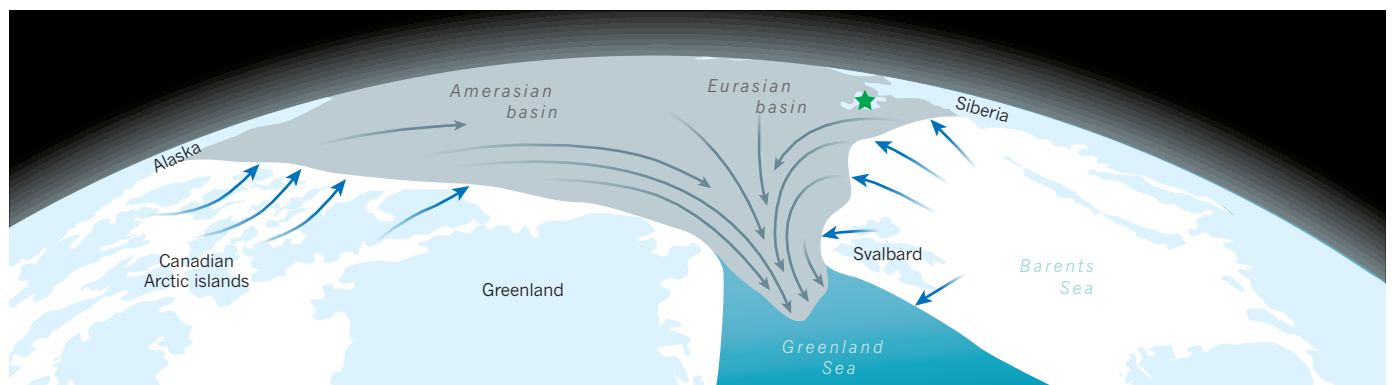
Ice cover across the Arctic Ocean has commonly been viewed as floating sea ice, formed as a relatively thin veneer from the freezing of seawater. Much effort has been directed towards understanding the behaviour of this form of ice cover and its recent precipitous decline<sup>1</sup>. But there has been a long-standing debate about a different kind of ice — ice of a thicker, more terrestrial origin, that might have existed across the Arctic basin in the distant past<sup>2</sup>. Writing in *Nature Communications*, Jakobsson *et al.*<sup>3</sup> bring this ancient ice shelf to life from a set of remarkable profiles

of elevated ridges that stand above the deep basins of the Arctic Ocean.

Early images<sup>4</sup> of portions of the Arctic sea floor revealed parallel grooves, scours and rumples — the telltale signs of ice having contact with and ploughing the sea floor. But the small number of features and their large geographic separation suggested to most scientists that floating and drifting icebergs were to blame<sup>5,6</sup>. Jakobsson and colleagues demonstrate that the ploughing pattern observed on widely separated ridges reflects a consistent and uniform direction of ice movement across the entire basin, as others have also reported<sup>7</sup>.

This key set of observations can be explained only by the presence of a single, widespread ice shelf (Fig. 1), one that flowed freely between large ridges, but became grounded and left its trail across the shallower regions of the Arctic basin. The ice shelf — estimated at some 4.1 million square kilometres on the basis of the numbers given in Jakobsson and colleagues' paper — was four times the area of all the ice shelves now found in Antarctica, including the immense Ross and Ronne-Filchner ice shelves. It must have been about 1 kilometre thick, allowing it to float over the great depths of the Amerasian and Eurasian basins, but touching down and ploughing across the shallower regions.

By tracing the glacial furrows towards their points of origin, Jakobsson and co-workers conclude that land-based ice sheets must have fed the ice shelf from three separate centres: the eastern Siberian lowlands; the Scandinavian regions of the Barents Sea; and high Arctic islands in north central Canada (Fig. 1). Furthermore, the authors found evidence — in the form of microfossils and of magnetic properties in sedimentary cores collected from some



**Figure 1 | Sketch of an ancient ice shelf across the Arctic Ocean.** Jakobsson *et al.*<sup>3</sup> report evidence of a one-kilometre-thick ice shelf (grey region) that covered the entire Arctic basin during the penultimate glaciation. Blue arrows indicate feeding ice streams from the Canadian Arctic islands, Siberia and Svalbard; grey arrows represent flow lines of the ice shelf. Modern locations of Greenland, Svalbard, Alaska and Siberia are shown in pale blue; green asterisk indicates the location of the New Siberian Islands. The rough locations of the Amerasian and Eurasian basins beneath the Arctic Ocean are also indicated. (Figure adapted from ref. 3.)

of the ridges — that this Great Arctic Ice Shelf (GAIS) existed not during the most recent glaciation (the Wisconsinan, about 22,000 to 18,000 years ago), but instead about 160,000 to 140,000 years ago, during the penultimate glaciation.

The antiquity of the GAIS helps to explain why some evidence is missing for the ice sheets that fed it. The regions from which the ice was derived are now shallow, broad shelves and coastal plains of Canada, Russia and north-western Europe. During the Wisconsinan and the subsequent sea-level rise, these regions were exposed to erosion and sedimentation, which may have masked the direct footprint of the ice sheets<sup>2</sup>. Even so, some regions that are clearly in the path of the ice sheets that covered eastern Russia have yielded no evidence of glaciation at all<sup>8</sup>. These areas include the remote New Siberian Islands, which may not have yet undergone complete geological examination<sup>9</sup>.

An ice shelf of the extent suggested by Jakobsson and colleagues' work is very different from the expansive sea ice that covers the Arctic Ocean today. Sea ice constantly opens and closes, providing gaps for light; this allows photosynthesis to occur and thus to fuel the vast Arctic marine ecosystem. Just as scientists contemplate the effect of more open water as Arctic sea ice is progressively reduced in response to climate change, an equally daunting task is to imagine how the ecosystem could function under the continual darkness and cover that the GAIS would have caused. A more limiting scenario for life could not be imagined, with the exception of the subglacial lakes that lie beneath the Antarctic ice sheet.

Equally important is the role that the GAIS must have had in regulating ocean chemistry while not greatly changing sea level. The implications of the changes in ocean chemistry for estimates of ancient sea levels can be appreciated by considering how climate scientists conventionally reconstruct sea-level oscillations. To reconstruct ancient glaciations largely lacking geological features that directly indicate shorelines, scientists have used the geochemical history of the ocean recorded in the shells of marine plankton called foraminifera. Glacial oceans are enriched in heavy-oxygen atoms (<sup>18</sup>O), and so the change in chemistry over time — measured as the ratio of light-oxygen atoms to heavy oxygen in foraminifera shells — is used to gauge sea-level changes caused by the comings and goings of large ice sheets<sup>10</sup>.

This method has been validated for the most recent glacial cycle, because direct measures of sea level are an exceedingly good match to estimates based on ice-sheet volume and changes in ocean chemistry<sup>11</sup>. But estimates for sea-level change during the penultimate glaciation have always been problematic, because the few direct shoreline records of sea level are higher than estimates based on the oxygen-isotope record corresponding to that period<sup>12</sup>. Jakobsson and co-workers' reconstructed

GAIS seems to explain this mismatch, and reminds climate scientists that they should check their assumptions as they continue to account for glacial ice and its effects on sea level during the numerous glacial episodes of the past 2.5 million years. ■

**Eugene Domack** is at the College of Marine Science, University of South Florida, St. Petersburg, Florida 33701, USA.  
e-mail: edomack@usf.edu

1. Overland, J. E. & Wang, M. *Geophys. Res. Lett.* **40**, 2097–2101 (2013).
2. Niessen, F. *et al. Nature Geosci.* **6**, 842–846 (2013).
3. Jakobsson, M. *et al. Nature Commun.* **7**, 10365 (2016).

4. Polyak, L., Edwards, M. H., Coakley, B. J. & Jakobsson, M. *Nature* **410**, 453–457 (2001).
5. Kristofferson, Y. *et al. Paleoceanography* **19**, PA3006 (2004).
6. Dowdeswell, J. A. *et al. Quat. Sci. Rev.* **29**, 3518–3531 (2010).
7. Engels, J. L., Edwards, M. H., Polyak, L. & Johnson, P. D. *Earth Surf. Processes Landforms* **33**, 1047–1063 (2008).
8. Andreev, A. A. *et al. Boreas* **33**, 319–348 (2004).
9. Basilian, A. C., Nikol'skiy, P. A. & Anisimov, M. A. *IPY News* **12**, 7–9 (2008).
10. Chappell, J. & Shackleton, N. J. *Nature* **324**, 137–140 (1986).
11. Dutton, A. *et al. Science* **349**, aaa4019 (2015).
12. Johnson, R. G. & Andrews, J. T. *Quat. Res.* **12**, 119–134 (1979).

This article was published online on 3 February 2016.

## AGEING

# Out with the old

**The selective elimination of cells that have adopted an irreversible, senescent state has now been shown to extend the lifespan of mice and to ameliorate some age-related disease processes. [SEE ARTICLE P.184](#)**

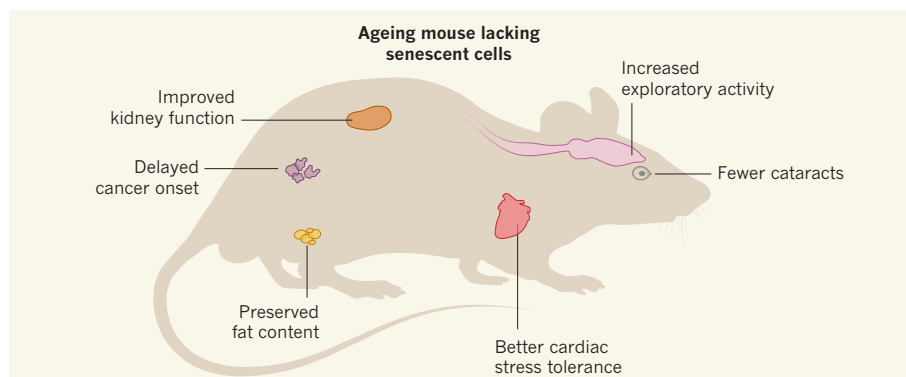
JESÚS GIL & DOMINIC J. WITHERS

**T**he ability to fight the ageing process has been a long-held human desire. Although this quest often seems to be driven by vanity, ageing is the major risk factor for many of the diseases that plague modern society. More than 50 years ago, it was suggested that ageing is linked to a state of arrested cell growth known as senescence<sup>1</sup>, but this link has remained unproven, and the molecular basis for organismal ageing has been elusive. In this issue, Baker *et al.*<sup>2</sup> (page 184) demonstrate that the removal of senescent cells does indeed delay ageing and increase healthy lifespan (healthspan).

Senescence is a cellular state in which cells permanently stop dividing. It is mediated by two signalling pathways — the

*p53* pathway and the *p16<sup>Ink4a</sup>–Rb* pathway. Senescent cells secrete a complex cocktail of factors called the senescence-associated secretory phenotype (SASP), which includes matrix metalloproteinases (enzymes that break down the extracellular matrix) and pro-inflammatory signalling molecules. Such cells have been shown to accumulate during ageing, and their presence has been associated with a broad range of diseases, including diabetes, kidney disease and many cancers<sup>3</sup>.

The group that performed the current study previously showed that removing senescent cells from a mouse model of accelerated ageing delays the onset of several disease-related processes<sup>4</sup>. However, the relevance of these observations to the normal ageing process was unclear. Baker *et al.* have now directly tackled this uncertainty using a genetically engineered



**Figure 1 | Improving healthspan.** Senescent cells, which are in a state of irreversible growth arrest, accumulate in various organs during ageing and are associated with age-related diseases in many tissues. Baker *et al.*<sup>2</sup> selectively eliminated senescent cells in ageing mice. This increased healthy lifespan, reducing many age-related, disease-associated abnormalities.



mouse model that they had developed previously<sup>4</sup>, called *INK-ATTAC*. These mice produce a caspase enzyme specifically in cells that express the *p16<sup>INK4a</sup>* gene. The caspase can be activated by the injection of a drug; the activated caspase then triggers cell death, eliminating senescent cells in which it is expressed.

Baker and colleagues found that the elimination of *p16<sup>INK4a</sup>*-expressing cells increased lifespan, regardless of the sex or strain of mouse examined, and ameliorated a range of age-dependent, disease-related abnormalities, including kidney dysfunction and abnormalities in heart and fat tissue (Fig. 1). The authors observed increased activity and exploratory behaviour and a decrease in the incidence of cataracts (although this improvement was strain-dependent). Senescent-cell removal also delayed the onset of cancer, without affecting the range of observed tumour types. Together, these findings suggest that the accumulation of *p16<sup>INK4a</sup>*-expressing cells during normal ageing shortens healthspan.

The *INK-ATTAC* mouse is a powerful model with which to investigate the physiological relevance of senescence, but it is not without limitations. For instance, the model is assumed to selectively eliminate senescent cells — and although not all *p16<sup>INK4a</sup>*-expressing cells are necessarily senescent, the *ATTAC* transgene that produces the caspase seems to be expressed only in senescent cells. However, it could be that drug treatment kills only ‘late senescence’ cells<sup>5</sup>, which express high levels of *p16<sup>INK4a</sup>* and *ATTAC*, rather than triggering a more general elimination of senescence. Moreover, drug treatment does not kill some senescent cells, including immune cells called lymphocytes as well as liver and colon cells, which limits the reach of the model. An improved characterization of the cell types that are eliminated is needed to fully understand the basis of the extended healthspan of these animals.

Another caveat is that the inducible elimination of senescent cells requires twice-weekly, long-term injections into the abdomen. Males that were injected with a control solution rather than the drug typically had shorter lifespans than normal mice, perhaps because of this intensive treatment regime. More-sophisticated model animals, in which senescent cells can be ablated in different tissues at different times and without the need for repeated injections, would help to extend the current findings.

Although the ablation of senescent cells ameliorates some age-related defects, it has no effect on others, including declines in motor performance, muscle strength and memory. This could reflect limitations of the *ATTAC* model. However, it might also suggest that senescent cells are involved in the progression of only some diseases.

Why might eliminating just the few cells that are senescent have beneficial effects

in a range of tissues? Baker and colleagues’ analysis of the kidney might help to explain this observation and clarify why senescent cells can be so disruptive during ageing. A striking disease-associated change often arises in aged kidneys, in capillary networks called glomeruli. However, the authors observed senescence primarily in another cell type, the epithelial cells of the kidney tubules. This suggests that SASP components secreted by epithelial cells could be responsible for disease in the glomeruli.

A search for compounds that can selectively eliminate senescent cells is under way<sup>6,7</sup>, and could be an important step in translating the findings of Baker and colleagues’ study to combating diseases of ageing in humans. An alternative therapeutic approach could be to repress the SASP. Indeed, inhibition of JAK proteins, which mediate the actions of some cytokines (a type of signalling molecule), reduces the SASP and alleviates frailty in old mice<sup>8</sup>. Rapamycin, a drug that is used as an immunosuppressant in humans, also robustly extends mouse lifespan<sup>9</sup> and regulates the SASP<sup>10,11</sup>. Thus, common therapeutic mechanisms acting on the SASP might underlie the beneficial effects of both rapamycin and senescent-cell ablation on lifespan and healthspan.

It is worth noting that senescence is a protective response that limits tissue scarring

(fibrosis) and cancer. Cells that express senescence markers are also involved in wound healing. Interestingly, the current study suggests that, although ablating senescent cells impairs wound healing, in general it has limited negative effects, and the authors found no evidence for increased fibrosis or cancer development. Nonetheless, any future senescence-based therapies must take care to control for possible detrimental consequences. ■

**Jesús Gil and Dominic J. Withers** are at the MRC Clinical Sciences Centre, Imperial College London, Hammersmith Campus, London W12 0NN, UK.

e-mail: d.withers@imperial.ac.uk

1. Hayflick, L. *Exp. Cell Res.* **37**, 614–636 (1965).
2. Baker, D. J. et al. *Nature* **530**, 184–189 (2016).
3. Muñoz-Espin, D. & Serrano, M. *Nature Rev. Mol. Cell Biol.* **15**, 482–496 (2014).
4. Baker, D. J. et al. *Nature* **479**, 232–236 (2011).
5. van Deursen, J. M. *Nature* **509**, 439–446 (2014).
6. Zhu, Y. et al. *Aging Cell* **14**, 644–658 (2015).
7. Chang, J. et al. *Nature Med.* **22**, 78–83 (2016).
8. Xu, M. et al. *Proc. Natl Acad. Sci. USA* **112**, E6301–E6310 (2015).
9. Harrison, D. E. et al. *Nature* **460**, 392–395 (2009).
10. Herranz, N. et al. *Nature Cell Biol.* **17**, 1205–1217 (2015).
11. Laberge, R. M. et al. *Nature Cell Biol.* **17**, 1049–1061 (2015).

This article was published online on 3 February 2016.

#### PLASMA PHYSICS

## Compact coupling for a two-stage accelerator

**Injecting electrons that have been accelerated by a laser-powered plasma wave into a second plasma wave represents a two-step electron accelerator. With 100 such steps, collider applications become possible. SEE LETTER P.190**

**BRIGITTE CROS**

**L**aser-plasma acceleration is a means of accelerating electrons in a plasma — a fluid of positively and negatively charged particles — using electric fields that are 1,000 times stronger than those of conventional particle accelerators operating in a vacuum. However, despite major advances in the past few years, reliable laser-plasma accelerators have still not been built. On page 190 of this issue, Steinke *et al.*<sup>1</sup> describe a way in which the coupling of electron beams can be controlled with plasmas. The authors’ achievement could lead to the development of laser-driven electron accelerators that can be scaled up to achieve energy gains on a par with those of today’s particle accelerators. Such electron accelerators would fit on a tabletop,

and could replace conventional machines at medical facilities, for instance.

Plasma waves — collective oscillations of electrons on a background of heavier, stationary ions — are intrinsic characteristics of plasmas. The oscillating electrons lead to a charge separation in the plasma wave that is associated with an electric field of up to several hundred gigavolts per metre. Plasma waves can be generated by the application of an intense, ultrashort laser pulse<sup>2</sup> (Fig. 1), which acts like a snow plough to expel electrons from the most intense regions of the plasma.

Depending on the laser intensity, plasma waves exist in either the quasilinear or the nonlinear regime (according to the wave equations that describe them). In the quasilinear case, plasma waves provide a periodic accelerating structure for injected particle beams

(Fig. 1). Alternatively, nonlinear interactions between the laser pulse and the plasma electrons during laser propagation cause the plasma waves to form spherical cavities<sup>3</sup> that can trap and accelerate a fraction of the plasma electrons. For example, electrons in plasmas are typically accelerated to energies in the 100-megaelectronvolt range across millimetre distances, for laser systems that have a maximum power of 30 terawatts (1 terawatt is  $10^{12}$  W). With the development of petawatt-class ( $10^{15}$  W) laser systems, acceleration in the nonlinear regime — extended to several centimetres by laser guiding — has generated multi-gigaelectronvolt electron bunches<sup>4</sup>. (For comparison, the Stanford Linear Accelerator Center in Menlo Park, California, produces electrons of 50 GeV over 3 km, whereas a plasma can achieve<sup>4</sup> 5 GeV in about 10 cm.)

To further increase the accelerated electron energy, a large electric field must be maintained over the longest possible distance (up to metres for the quasilinear regime). Lowering the density of the electrons in the plasma helps to reduce nonlinear effects and to increase the acceleration distance, provided that the laser intensity is maintained at the required level over the whole distance. This can be achieved by guiding the laser beam externally, using either a preformed plasma structure or a gas-filled capillary tube. However, the interaction between the laser and the plasma causes a transfer of the laser beam's energy to the plasma, leading to 'laser depletion'. This problem, and the limitation on acceleration distance, can be resolved by using several plasma stages in

sequence. Supplying a fresh laser beam for each subsequent accelerating stage makes the process scalable in electron energy and repetition rate, with a reasonable amount (10–100 joules) of laser energy per beam.

Steinke *et al.* have combined an ensemble of state-of-the-art plasma designs to achieve what could potentially be developed into an all-plasma compact accelerator driven by lasers. For the first stage, they use the nonlinear regime to create the electron beam from a gas jet; the nonlinear regime is necessary because, in the quasilinear regime, no electrons from the plasma itself can be trapped and accelerated. The authors use the quasilinear regime for the second stage, at which the accelerated electrons from the first stage, having energies of 100 MeV, are injected. Notably, the authors' components for coupling the injected electrons with the plasma wave and for focusing the laser have been reduced from the metre scale to the centimetre scale, with their entire set-up being on a scale of 10 cm.

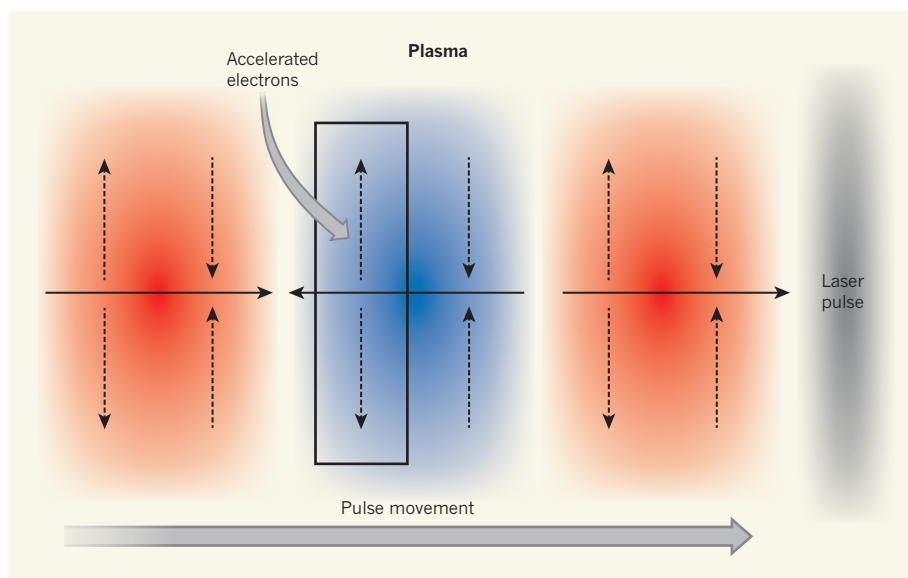
The novelty of their scheme relies on the use of plasmas for every key element of the accelerator: the electron source is produced when a laser pulse hits a jet of gas, which becomes ionized as a plasma and leads to an electron beam of short duration (typically 10 femtoseconds; 1 femtosecond is  $10^{-15}$  s); the electron beam is focused by a plasma lens that is produced by a column of plasma created by discharge inside a capillary (the magnetic field of the plasma lens symmetrically confines the electrons); and the electron beam is accelerated in a plasma wave that is driven by a pulsed laser beam focused

with a plasma mirror (a mirror that reflects light above a threshold intensity at which the surface of the material is ionized). In contrast to previous experiments, two independent laser beams are used, driving two separate accelerating plasma stages. The consequence of using plasmas for each element is to maintain the overall compactness of the beam-line, thus preserving the global accelerating electric field.

Although this experiment is a remarkable achievement, room for improvement remains. The electron beam was externally injected into the plasma wave and accelerated with a maximum energy gain of 100 MeV and a 3.5% charge-coupling efficiency (the fraction of charge at the entrance of the second stage that will be accelerated at the exit). Steinke and colleagues present simulations showing that the trapping efficiency of electrons in the plasma wave might be increased to 90% by, for example, injecting an electron beam with optimized proportions specifically into the accelerating and focusing regions of the plasma wave. This could be achieved by using an electron beam (in which the spread of the electrons' energy is smaller than that in Steinke and colleagues' experiments) that is focused on a spot of typically less than  $10\ \mu\text{m}$ , and coupling this with a plasma wave that is larger than the electron beam and that is excited by a laser with a uniform amplitude (guided in a way that prevents oscillations). An energy gain in the 10-GeV range would be expected for metre-scale plasmas, although plasma waveguides of this size are not currently available. In addition to the quality of the coupling of the electron bunch as it enters the plasma, its stability in the plasma wave must be controlled precisely in space and time as it is accelerated, to achieve a high-quality electron beam.

Although there is a long way to go before a laser-plasma accelerator with 100 stages can be built for collider applications, intermediate-energy accelerators in the 10-GeV range are a realistic possibility in the near future, as test facilities for accelerator physics and other applications. Numerous research and development activities are under way worldwide, including collective efforts<sup>5</sup> in Europe to design an intermediate-energy laser-plasma accelerator. Such activities hold the promise of shrinking the size and cost of future machines. ■

**Brigitte Cros** is at the *Laboratoire de Physique des Gaz et des Plasmas, CNRS, Université Paris-Sud, Université Paris Saclay, 91405 Orsay, France.*  
e-mail: [brigitte.cros@u-psud.fr](mailto:brigitte.cros@u-psud.fr)



**Figure 1 | Structure of a plasma wave.** An initial laser pulse propagating through a plasma displaces electrons (not shown) in the plasma. These electrons can generate a plasma wave in the 'quasilinear' regime, in which the wave has oscillating electric fields (longitudinal component shown by solid black arrows, transverse component by dotted black arrows). Injected electrons in tune with the accelerating phase (blue; decelerating phases in red) of the longitudinal wave and with the focusing fields pointing away from the horizontal axis — that is, within the black oblong — will be focused and accelerated. Steinke and co-authors<sup>1</sup> show that such a plasma wave can accelerate an electron beam produced from a first-stage plasma accelerator. In principle, using more than two stages will further scale the energy gain.

1. Steinke, S. *et al.* *Nature* **530**, 190–193 (2016).
2. Tajima, T. & Dawson, J. M. *Phys. Rev. Lett.* **43**, 267–270 (1979).
3. Pukhov, A. & Meyer-ter-Vehn, J. *Appl. Phys. B* **74**, 355–361 (2002).
4. Leemans, W. P. *et al.* *Phys. Rev. Lett.* **113**, 245002 (2014).
5. [www.eupraxia-project.eu](http://www.eupraxia-project.eu)



## EBOLA

## Sequencing on the ground

The West African Ebola epidemic that began in 2014, and still threatens to re-emerge, prompted a flurry of efforts to sequence the genome of the virus as it spread. Genomic information can contribute to epidemiological tracking, identify mutations that increase the disease's virulence and inform therapeutic strategies. But sequences must be obtained fast if they are to help to control a raging epidemic.

On page 228 of this issue, Quick *et al.* describe a portable, real-time sequencing system that they used in local, resource-poor sites in Guinea during 2015 (J. Quick *et al.* *Nature* **530**, 228–232; 2016). The requisite equipment, including the tiny MinION genome sequencer (pictured), fitted into 50 kilograms of standard airline luggage. The researchers generated sequence data within 24 hours of receiving a patient's viral RNA sample — days to weeks less than it would have taken to receive data from a central sequencing facility. [Marian Turner](#)



## GENETICS

## Asymmetric breaks in DNA cause sterility

The part that the mouse gene *Prdm9* plays in generating double-strand breaks in DNA has now been linked to its putative role in speciation, thanks to experiments that use a 'humanized' version of the gene. [SEE ARTICLE P.171](#)

JIRI FOREJT

During the specialized meiotic cell divisions that produce mammalian eggs and sperm, an individual's two sets of chromosomes are shuffled in a process called meiotic recombination, producing genetically unique chromosomes to pass on to offspring. Recombination tends to occur in DNA regions known as hotspots, the locations of which are determined by a single DNA-binding protein, PRDM9. The presence of different versions (alleles) of the *Prdm9* gene in two subspecies of mice renders the hybrid offspring of these subspecies sterile<sup>1</sup>, but how the gene exerts this effect has been unknown. On page 171 of this issue, Davies *et al.*<sup>2</sup> link these two roles of *Prdm9*. They find that it is the rapid subspecies-specific degradation

of *Prdm9* DNA-binding sites that underlies sterility in mouse hybrids.

The male offspring of crosses between the mouse subspecies *Mus musculus musculus* and *Mus musculus domesticus* are sterile. It has been proposed that sterility arises for two reasons. First, the chromosomes from the two subspecies do not recognize one another normally because their DNA sequences have diverged, so they do not align correctly — to use the technical terminology, they do not synapse. Such asynapsis results in the death of sperm-cell progenitors, although the mechanism underlying this process is not well understood. Second, transcription from the sex chromosomes is not inhibited as it is in the meiosis of healthy testes<sup>3</sup>.

Both of these defects can prevent normal sperm development. The extent of the damage

is regulated by interaction between the different alleles of *Prdm9* and another gene, *Hstx2*, that are present in each subspecies<sup>1,4</sup>. But an understanding of which diverging sequences prevent synapsis, and of the molecular mechanisms underlying the roles of *Prdm9* and *Hstx2*, has remained out of reach.

Davies *et al.* now report considerable progress in defining the role of *Prdm9* in hybrid sterility. They focused on a structure called the Cys<sub>2</sub>His<sub>2</sub> zinc-finger domain of mouse PRDM9, which is responsible for the protein's ability to bind to specific DNA sequences. DNA binding enables PRDM9 to add methyl groups to a lysine amino-acid residue on the protein histone H3, around which DNA is packaged in the nucleus. Such methylation is a prerequisite for the production of double-strand breaks (DSBs) in nearby DNA; these breaks are subsequently repaired through meiotic recombination, with the unbroken equivalent chromosome (the homologous chromosome) used as a genetic template for repair<sup>5</sup>.

The authors engineered a 'humanized' form of *Prdm9*, in which they replaced the zinc-finger domain from *M. m. domesticus* *Prdm9* with the equivalent sequence from the human gene, which binds to a different DNA sequence. The humanized male mice were fertile but displayed a pattern of recombination hotspots that shared little overlap with the hotspots of wild-type *M. m. domesticus* mice. Davies and colleagues then crossed the humanized males to

*M. m. musculus* females. The hybrids were fertile, unlike hybrids from wild-type *M. m. domesticus* males, and the development of their sperm was normal. This finding clearly shows that it is the zinc-finger domain of PRDM9 that is responsible for sterility in hybrid mice.

Analysis of DSB sites in the sperm-cell progenitors of infertile hybrids revealed that most DSBs occurred in asymmetric hotspots — that is, the PRDM9 protein produced from the set of chromosomes from one parent bound preferentially to sites on the homologous chromosome. This asymmetry arises because of an evolutionary quirk known as the hotspot paradox. In this paradox, if any mutation that arises in a hotspot prevents PRDM9 from binding to that site, it is preferentially transmitted to the next generation, because the mutated sequence will be used as a template to repair DSBs in the unmutated hotspot on the homologous chromosome. In this way, PRDM9 binding sites are gradually lost over time<sup>6</sup>. The asymmetry develops because these sites are not lost from closely related subspecies, in which PRDM9 binds a different sequence.

Because humans diverged from mice long before the two mouse subspecies diverged, the

hotspots to which the humanized *Prdm9* bound were mostly symmetrical; that is, the protein bound to the same sites on both sets of mouse chromosomes. The authors conclude that the fertility of the humanized mice is associated with this rescue of hotspot symmetry. In support of this assertion,

***It is the zinc-finger domain of the PRDM9 protein that is responsible for sterility in hybrid mice.***

the degree of asymmetric PRDM9 binding varied between chromosomes in the sterile hybrids, and this correlated with the previously reported<sup>3</sup> frequency with which meiotic asynapsis occurred in each chromosome pair. Thus, reduced PRDM9 binding symmetry is linked to impaired DSB repair and predisposition to asynapsis.

Davies and colleagues' results provide considerable insight into how hybrid sterility arises in this particular model. However, hybrid sterility as a mechanism to prevent gene exchange between related species is a widespread phenomenon throughout the animal and plant kingdoms<sup>7</sup>, and the general significance of *Prdm9* or *Prdm9*-like genes for the process of reproductive isolation is still

uncertain. The next step will be to extend the authors' work, both to analyse some of the other *Prdm9* alleles that have been identified in various mouse subspecies<sup>8,9</sup> and to consider other mammalian models of hybrid sterility. Only then will we be able to evaluate the fascinating possibility that *Prdm9* has a recurrent role in reproductive isolation between closely related mammalian species<sup>10</sup>. ■

**Jiri Forejt** is at the Division BIOCEV, Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, 142 20 Prague 4, Czech Republic.  
e-mail: jiri.forejt@img.cas.cz

1. Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J. C. & Forejt, J. *Science* **323**, 373–375 (2009).
2. Davies, B. *et al.* *Nature* **530**, 171–176 (2016).
3. Bhattacharyya, T. *et al.* *Proc. Natl Acad. Sci. USA* **110**, e468–e477 (2013).
4. Bhattacharyya, T. *et al.* *PLoS Genet.* **10**, e1004088 (2014).
5. Baudat, F., Imai, Y. & de Massy, B. *Nature Rev. Genet.* **14**, 794–806 (2013).
6. Myers, S. *et al.* *Science* **327**, 876–879 (2010).
7. Maheshwari, S. & Barbash, D. A. *Annu. Rev. Genet.* **45**, 331–355 (2011).
8. Buard, J. *et al.* *PLoS ONE* **9**, e85021 (2014).
9. Kono, H. *et al.* *DNA Res.* **21**, 315–326 (2014).
10. Oliver, P. L. *et al.* *PLoS Genet.* **5**, e1000753 (2009).

This article was published online on 3 February 2016.

a vague pattern of light and shadowy regions, instead of a sharp diffraction pattern with distinct spots (see Figure 1 of the paper<sup>2</sup>). This vague picture constitutes a kind of 'continuous' diffraction<sup>2–4</sup> pattern, and contains information about the structure of the molecules in the crystals at resolutions extending well beyond the limit of Bragg diffraction.

Continuous diffraction has frequently been observed in diffraction patterns of protein crystals, but had not previously been used for structural determination. Ayer and co-workers' approach is similar to that used for single-molecule imaging<sup>5</sup> to obtain structural information from the continuous diffraction pattern, working on the basis that the diffraction arises from the sum of aligned single-molecule diffraction patterns.

The authors' method requires both Bragg and continuous diffraction data, and treats these diffractions separately. First, the Bragg diffractions were analysed using conventional techniques, generating an electron-density map of the target molecule at low resolution. The researchers used this map to generate a low-resolution outline of the 3D molecular structure. This, in turn, was used to constrain a higher-resolution 3D image of the electron density, generated from the continuous diffraction data using an 'iterative phasing' algorithm<sup>6,7</sup>. The structure was finally refined using a procedure similar to that used for single-particle cryo-electron microscopy data<sup>8</sup>.

Ayer *et al.* used this approach to extend the

## CRYSTALLOGRAPHY

# Resolution beyond the diffraction limit

**A method has been devised that extends the resolution of X-ray crystal structures beyond the diffraction limit. This might help to improve the visualization of structures of proteins that form 'poorly diffracting' crystals. SEE LETTER P.202**

JIAN-REN SHEN

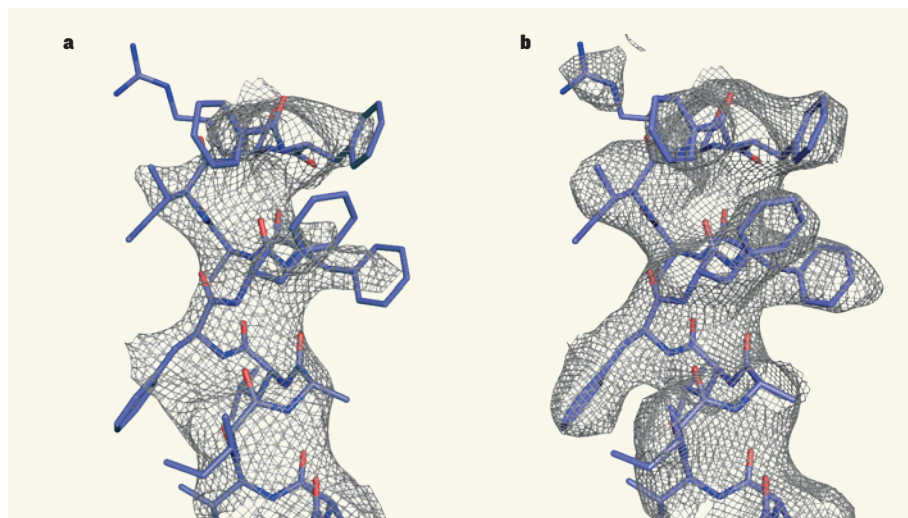
In 1913, Lawrence Bragg described the law that characterizes the diffraction of X-ray beams by crystalline materials<sup>1</sup>. Since then, crystallographers have used this law to solve the atomic structures of a wealth of compounds, ranging from simple table salt to extremely complex biomacromolecules. The resolution at which a structure can be visualized depends on the diffraction limit (or the Bragg limit), which in turn depends on the regular arrangement of molecules within crystalline materials. The degree of this regularity is often low in crystals of biomacromolecules, and this creates a major barrier to visualizing these molecules at the atomic level. On page 202 of this issue, Ayer *et al.*<sup>2</sup> report an approach to solving molecular structures at resolutions higher than the diffraction limit.

The X-ray diffraction defined by Bragg's law

is the sum total of diffraction patterns from molecules that strictly occupy the positions of a crystal's lattice units. Obtaining such a strictly ordered crystal often represents a big obstacle to the study of hydrophobic, membrane-spanning proteins and their complexes, limiting the resolution at which the structures of these biologically important molecules can be solved. The low degree of order in protein crystals can be caused by translational and rotational displacements of the protein molecules from their ideal positions, and by dynamic motions within the molecules.

Ayer and colleagues propose that molecules in crystalline materials can be considered as rigid units, and that the disorder is caused by translational displacements of these units from their crystal-lattice positions. These displacements cause scattered X-rays to deviate from the trajectories generated by a perfectly ordered crystal lattice, giving rise to





**Figure 1 | Electron-density maps of a section of photosystem II.** Ayer *et al.*<sup>2</sup> have developed a method that improves the resolution of X-ray crystal structures by analysing information from both conventional Bragg diffraction and continuous diffraction — which occurs when molecules in crystals deviate from their exact lattice positions only through translational displacements. **a**, Here, the structure of an  $\alpha$ -helix chain of photosystem II, a photosynthetic protein complex, is superimposed on the electron-density map (grey mesh) of the chain obtained by analysing only Bragg X-ray diffraction. The resolution is 4.5 Å. **b**, Using the authors' method, the resolution of the electron-density map is improved to 3.5 Å, and the molecular structure fits more precisely into the map.

resolution of the structure of photosystem II (PSII), an extremely large, dimeric membrane-protein complex (the molecular weight of the dimer is about 700 kilodaltons). PSII splits water into electrons, protons and oxygen during photosynthesis, thereby sustaining aerobic life on Earth by providing oxygen and converting light energy into biologically useful chemical energy. The structure of PSII has previously been solved at a resolution of 1.9 ångströms using X-rays from a conventional synchrotron source<sup>9</sup>, and at a resolution of 1.95 Å using femtosecond X-ray pulses (1 femtosecond is  $10^{-15}$  s) from a free-electron laser (XFEL)<sup>10</sup>. These structures indicated that the PSII molecules are rigid within the crystals, at least up to this level of resolution. However, some of Ayer and colleagues' PSII crystals diffracted to only 4.5 Å resolution, limiting the authors' ability to study the complex's reaction dynamics.

To obtain better resolution, the researchers collected a data set from a stream of micrometre-scale PSII crystals using powerful femtosecond XFEL pulses, and thus obtained an X-ray structure of 4.5 Å based on Bragg diffractions. They then refined this to 3.5 Å resolution using their method based on continuous diffraction data. The resulting electron-density map showed more-detailed features of many amino-acid side chains (Fig. 1) and cofactors associated with the large protein complex, compared with the structure derived from Bragg diffraction data alone. This great improvement hints at the potential of this approach for extending resolution limited by Bragg diffraction.

The extent to which rotational displacements

and dynamic motions in crystals can generally be ignored remains to be seen — this will determine how useful continuous diffraction data will be for improving resolutions. Nevertheless, the authors' approach might find broad applications, from structural analysis beyond Bragg diffractions to single-molecule imaging, and help to improve the accuracy of structures derived from X-ray data.

#### ASTROPHYSICS

## Exoplanets hidden in the gaps

**Planets develop from the disk of dust and gas that surrounds a newly formed star. Observations of gaps in the disks of four such systems have allowed us to start unravelling the processes by which planets form.**

PAUL HO

**T**he brightest objects in the night sky, besides the Moon, are the inner planets of the Solar System: because they are close to us, the light they reflect from the Sun is brighter than that of the stars. But exoplanets are as distant as stars, and are therefore much dimmer, making them difficult to detect directly even with the biggest telescopes. Moreover, their companion stars are usually a billion times brighter than they are, and blurring and broadening of the stellar image

In particular, the technique is potentially a great step forward for those seeking high-resolution structural information for many 'poorly diffracting' protein crystals and their complexes.

It will be interesting to see how much further resolution can be improved by incorporating continuous diffraction data with X-ray data collected from both XFEL and synchrotron sources. Meanwhile, the reported iterative-phasing approach may allow structures to be solved without the use of additional methods to retrieve phase information from diffraction data. It is also to be hoped that the reported iterative-phasing algorithms will be made user-friendly and widely available. ■

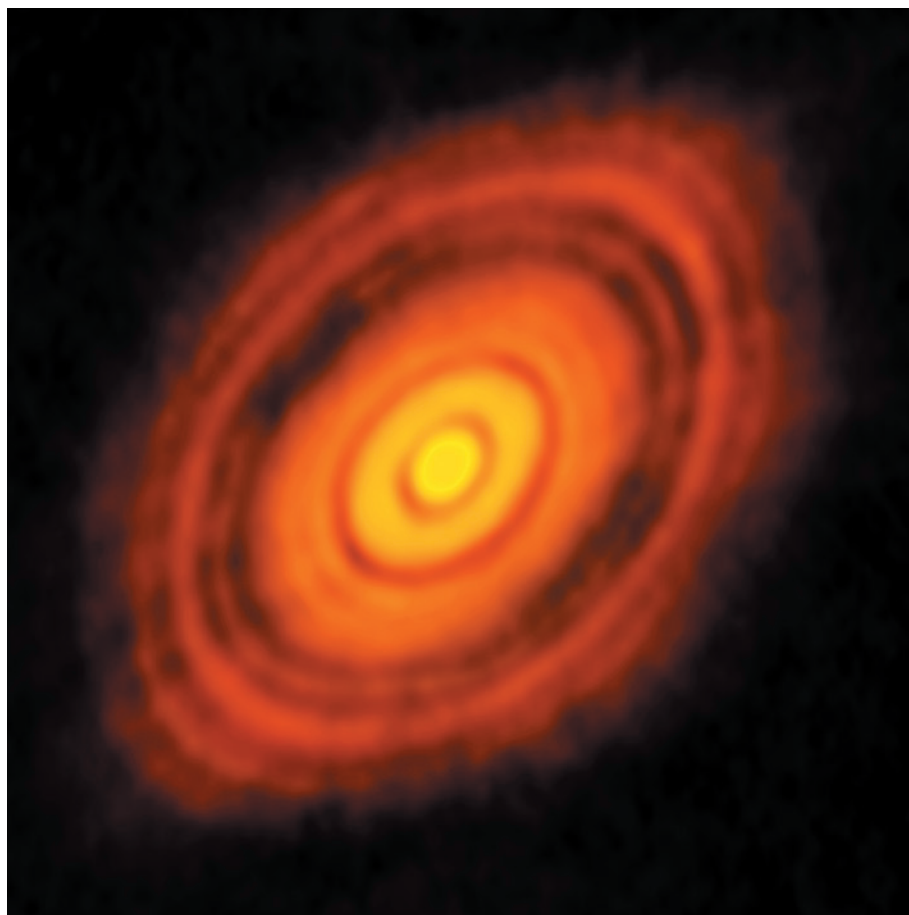
**Jian-Ren Shen** is at the Photosynthesis Research Center, Graduate School of Natural Science and Technology, Okayama University, Tsushima Naka 3-1-1, Okayama 700-8530, Japan.

e-mail: shen@cc.okayama-u.ac.jp

1. Bragg, W. L. *Proc. Camb. Phil. Soc.* **17**, 43–57 (1913).
2. Ayer, K. *et al. Nature* **530**, 202–206 (2016).
3. Pérez, J., Faure, P. & Benoit, J.-P. *Acta Crystallogr. D* **52**, 722–729 (1996).
4. Welberry, T. R., Heerdegen, A. P., Goldstone, D. C. & Taylor, I. A. *Acta Crystallogr. B* **67**, 516–524 (2011).
5. Miao, J., Ishikawa, T., Robinson, I. K. & Murnane, M. M. *Science* **348**, 530–535 (2015).
6. Elser, V. J. *Opt. Soc. Am. A* **20**, 40–55 (2003).
7. Elser, V. & Millane, R. P. *Acta Crystallogr. A* **64**, 273–279 (2008).
8. Fischer, N. *et al. Nature* **520**, 567–570 (2015).
9. Urmenia, Y., Kawakami, K., Shen, J.-R., & Kamiya, N. *Nature* **473**, 55–60 (2011).
10. Suga, M. *et al. Nature* **517**, 99–103 (2015).

by Earth's atmosphere typically hides the nearby exoplanets. Writing in *Astronomy & Astrophysics*, van der Marel *et al.*<sup>1</sup> report that they have overcome these problems to acquire images of protoplanetary disks — the dust and gas surrounding a newborn star, from which planets form. Their observations show that nascent exoplanets have apparently begun to clear disk material from the inner zones of their systems.

To understand how planets form, we need to find systems in which the protoplanetary disk is detectable so that the relationship



**Figure 1 | The young star HL Tauri.** This image reveals that rings and gaps are present in the disk of dust and gas surrounding the star. The gaps were probably opened and maintained by unseen planets. Van der Marel *et al.*<sup>1</sup> have detected gas in the central cavities of four other protoplanetary disks.

between the disk's evolution and planetary formation can be studied. The interaction between a planet and its surrounding disk can provide clues. The best model for such systems may be the planetary rings around Saturn or Jupiter. The gaps between rings are maintained by 'shepherd' moons in the gaps: gravitational interactions and the exchange of angular momentum between a moon and the disk opens a gap and preserves it. The gaps are much easier to detect than the shepherd moons, because the gaps have a considerably larger area, and because there is a large contrast between the gap and the rings.

In 2015, a protoplanetary disk around the star HL Tauri was imaged<sup>2</sup> by the Atacama Large Millimeter/submillimeter Array (ALMA) — the largest ground-based observatory ever built, with a radio telescope that can detect the cold dust in disks with great sensitivity. The images showed that rings and gaps are present in the disk around HL Tauri (Fig. 1), just as they are around Saturn. Until now, this provided the best evidence of a system in which multiple planetary bodies had formed in a protoplanetary disk. Van der Marel and colleagues have now used ALMA to observe a sample of 'transitional' disks, which form around solar-type stars and have a

central cavity in which growing planets might be starting to clear the protoplanetary material.

The authors used ALMA's exceptionally high angular resolution to resolve the images, and used spectral lines associated with rare carbon isotopes in carbon monoxide (CO) to distinguish this gas in the disks. Locating the gas in disks is important for our understanding of planetary formation, because gas is typically 100 times more abundant than dust in the interstellar medium, and typically a few times less abundant than that in protoplanetary disks<sup>3</sup>. The authors observed the dust by studying its continuum emission (a large range of wavelengths, rather than just the narrow ranges of spectral lines), and made their observations at submillimetre wavelengths to suppress the interference from the companion star's light.

On the basis of their CO measurements, the authors find that the central cavity in the dust in each studied system is not totally empty, but contains gas. More specifically, they observe that the regions depleted of dust have radii ranging from 25 to 60 AU (1 AU is the distance from Earth to the Sun), whereas the regions depleted of gas have smaller radii (7 to 30 AU). Within the dust cavity (the region depleted of dust, but not of gas) there is at

least 10,000 times more gas than dust. The researchers propose that the depletion of gas and dust in the cavities could be an effect of heating by electromagnetic radiation from the central star, or of perturbation by unseen shepherd planets.

A planet's gravitational potential can open a gap in a disk by inducing orbital resonances — regular, periodic gravitational interactions with the disk material. But the viscosity of the gas in the disk allows transfer of the disk material's angular momentum away from the resonance points; this transfer enables the gas to refill the gaps. The interaction of gas with the planets can give rise to: inward migration of those planets; a balance of angular-momentum transport between the planets and the disk; and the continual accretion of material onto the planets.

The detailed physics of the exchange of angular momentum and energy between the disk and a planet is complicated. However, calculations<sup>4,5</sup> have shown that, with judicious choices of boundary conditions and disk viscosity, the relative sizes of the gas and dust cavities provide a means of measuring the mass of the possible planetary shepherd. By measuring the properties of the gas in the central cavities of the four observed protoplanetary disks, van der Marel and co-workers estimate that the typical mass of an unseen planet is about 10 times that of Jupiter. For comparison, the average disk mass of the observed systems is about 50 times that of Jupiter. Intriguingly, the dust cavities are usually about the size of Neptune's orbit. This implies that the planetary formation processes occurring in such systems might be similar to those that occurred in the Solar System.

The current study is just the beginning of an upcoming avalanche of ALMA studies. The ability to measure spectral lines, such as those of CO, is crucial because it allows the measurement of motions within gaseous disks. Van der Marel and colleagues' results clearly show rotation of the disks. More-sensitive data will allow us to see the motions and the structures of gas inflows. The results will begin to reveal how planets assemble in solar systems like our own. ■

**Paul Ho** is at the Academia Sinica Institute of Astronomy and Astrophysics, Taipei, 106 Taiwan.  
e-mail: pho@asiaa.sinica.edu.tw

1. van der Marel, N. *et al.* *Astron. Astrophys.* **585**, A58 (2016).
2. ALMA Partnership *et al.* *Astrophys. J.* **808**, L3 (2015).
3. Williams, J. P. & Best, W. M. J. *Astrophys. J.* **788**, 59 (2014).
4. Fung, J., Shi, J.-M. & Chiang, E. *Astrophys. J.* **782**, 88 (2014).
5. Dipierro, G. *et al.* *Mon. Not. R. Astron. Soc.* **453**, L73–L77 (2015).

**The author declares competing interests.** See [go.nature.com/lavkve](http://go.nature.com/lavkve) for details.



# Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice

Benjamin Davies<sup>1\*</sup>, Edouard Hatton<sup>1\*</sup>, Nicolas Altemose<sup>1,2</sup>, Julie G. Hussin<sup>1</sup>, Florencia Pratto<sup>3</sup>, Gang Zhang<sup>1</sup>, Anjali Gupta Hinch<sup>1</sup>, Daniela Moralli<sup>1</sup>, Daniel Biggs<sup>1</sup>, Rebeca Diaz<sup>1</sup>, Chris Preece<sup>1</sup>, Ran Li<sup>1,2</sup>, Emmanuelle Bitoun<sup>1</sup>, Kevin Brick<sup>3</sup>, Catherine M. Green<sup>1</sup>, R. Daniel Camerini-Otero<sup>3</sup>, Simon R. Myers<sup>1,2§</sup> & Peter Donnelly<sup>1,2§</sup>

**The DNA-binding protein PRDM9 directs positioning of the double-strand breaks (DSBs) that initiate meiotic recombination in mice and humans. *Prdm9* is the only mammalian speciation gene yet identified and is responsible for sterility phenotypes in male hybrids of certain mouse subspecies. To investigate PRDM9 binding and its role in fertility and meiotic recombination, we humanized the DNA-binding domain of PRDM9 in C57BL/6 mice. This change repositions DSB hotspots and completely restores fertility in male hybrids. Here we show that alteration of one *Prdm9* allele impacts the behaviour of DSBs controlled by the other allele at chromosome-wide scales. These effects correlate strongly with the degree to which each PRDM9 variant binds both homologues at the DSB sites it controls. Furthermore, higher genome-wide levels of such 'symmetric' PRDM9 binding associate with increasing fertility measures, and comparisons of individual hotspots suggest binding symmetry plays a downstream role in the recombination process. These findings reveal that subspecies-specific degradation of PRDM9 binding sites by meiotic drive, which steadily increases asymmetric PRDM9 binding, has impacts beyond simply changing hotspot positions, and strongly support a direct involvement in hybrid infertility. Because such meiotic drive occurs across mammals, PRDM9 may play a wider, yet transient, role in the early stages of speciation.**

In spite of its central role in evolution, the molecular mechanisms underlying speciation are not well understood. Only a small number of genes involved in speciation have been documented<sup>1</sup>, with only one such gene, *Prdm9*, known in mammals<sup>2,3</sup>. *Prdm9* contributes to hybrid sterility in male (PWD × B6)F1 mice from crosses between male *Mus musculus domesticus* C57BL/6 (hereafter B6) and female *Mus musculus musculus* PWD/Ph (hereafter PWD)<sup>4</sup>. Although its genetic basis is only partially understood<sup>5,6</sup>, this hybrid sterility is characterized by failure of pairing (synapsis) of homologous chromosomes and an arrested meiotic prophase owing to lack of repair of recombination intermediates<sup>2</sup>. Homologous recombination, and synapsis, are interdependent, essential meiotic processes<sup>7</sup>, and evidence suggests synapsis often nucleates at recombination sites<sup>8</sup>. Aside from the PWD × B6 cross, *Prdm9* allele and dosage have been associated with variation in measures of fertility and successful meiosis in many additional mouse crosses<sup>9</sup>.

PRDM9 has several functional domains, including a DNA-binding zinc-finger array, and a PR/SET domain responsible for histone H3 lysine 4 trimethylation (H3K4me3)<sup>10</sup> (Fig. 1a). By binding to specific DNA sequence targets, PRDM9 directs the positions of the DSB events that initiate meiotic recombination<sup>11</sup>. This results in these DSBs and downstream recombination events clustering into small discrete regions called hotspots<sup>12,13</sup>. The PRDM9 zinc-finger array, encoded by a minisatellite repeat, is highly polymorphic within and across mammalian species<sup>3,14–16</sup> and is among the fastest evolving regions in the genome, with strong evidence of natural selection influencing this evolution<sup>3</sup>. It is unknown whether PRDM9 zinc-finger array polymorphism has additional impacts, aside from direct alterations of DSB hotspot positions.

## Humanizing *Prdm9* restores hybrid fertility

To explore the DNA-binding characteristics of PRDM9, we generated a line of humanized B6 mice, by replacing the portion of mouse *Prdm9*

exon 10 encoding the zinc-finger array with the orthologous sequence from the human reference *PRDM9* allele (the 'B' allele) (Fig. 1a, Extended Data Fig. 1). A feature of PRDM9 (explored further below) is the co-evolution of its zinc-finger array with the genomic background in which it sits<sup>13,17</sup>. Minisatellite mutational processes at PRDM9 can produce new alleles with duplications, deletions or rearrangements within the zinc-finger array, yielding an almost complete change in PRDM9 binding sites, and thus hotspot locations<sup>14</sup>. Because the human PRDM9 zinc-finger array evolved on a lineage separated from mice for ~150 million years, our experimental approach allows assessment of functional properties of a PRDM9 zinc-finger array unaffected by changes it has induced in the background genome, similar to new alleles randomly arising in the population.

Humanization of the *Prdm9* zinc-finger array in B6 inbred mice had no effect on fertility (Extended Data Fig. 2) and cytogenetic comparisons revealed no significant impact on zygotene DSB counts (DMC1 immunoreactivity, Extended Data Fig. 2b), crossover counts (MLH1 foci, Extended Data Fig. 2c), normal sex body formation (γH2AX immunostaining, Fig. 1b) or quantitative measures of fertility and successful synapsis (see later). The full fertility of humanized mice implies there are unlikely to be any specific essential PRDM9 binding sites. One mechanism underlying speciation in many settings involves Dobzhansky–Muller incompatibilities: hybrid dysfunction arising from incorrect epistatic interactions<sup>1</sup>. Based on the above, it seems likely that if such interactions involving PRDM9 occur, they do not reflect constrained co-evolution of *Prdm9* with specific genes.

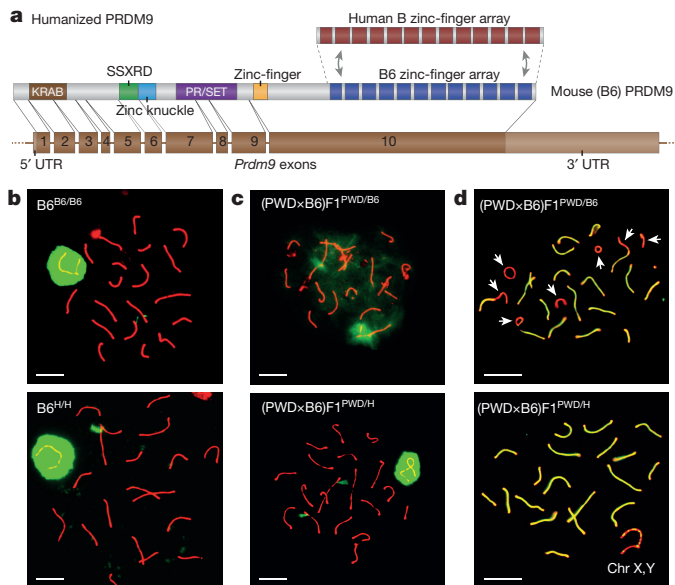
To explore the role of PRDM9 in fertility directly, we crossed PWD females with B6<sup>B6/H</sup> males. As expected<sup>18</sup>, male (PWD × B6)F1<sup>PWD/B6</sup> hybrids (we use superscripts to indicate *Prdm9* genotypes and write the female strain first in crosses) exhibited hybrid sterility as evidenced by failures in siring pups (Extended Data Fig. 2e), sex body formation

<sup>1</sup>The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, University of Oxford, Oxford OX3 7BN, UK. <sup>2</sup>Department of Statistics, University of Oxford, 24-29 St. Giles', Oxford OX1 3LB, UK.

<sup>3</sup>Genetics and Biochemistry Branch, National Institute of Diabetes, Digestive and Kidney Diseases, NIH, Bethesda, Maryland 20892, USA.

\*These authors contributed equally to this work.

§These authors jointly supervised this work.



**Figure 1 | Humanizing the zinc-finger domain of PRDM9 does not impact fertility.** **a**, Domain structure of the re-engineered PRDM9 protein **b**,  $\gamma$ H2AX staining of the sex body (green), SYCP3 staining of the chromosome axis (red) in late pachytene in  $B6^{B6/B6}$  (top) and  $B6^{H/H}$  (bottom). **c**, As **a**, but for  $(PWD \times B6)F1^{PWD/B6}$  and  $(PWD \times B6)F1^{PWD/H}$ . **d**, SYCP1 staining of the synaptonemal complex transverse filament (green), and SYCP3 staining of the chromosome axis (red) in pachytene for  $(PWD \times B6)F1^{PWD/B6}$  and  $(PWD \times B6)F1^{PWD/H}$ . Arrows, unsynapsed autosomes. Scale bars, 10  $\mu$ m.

(Fig. 1c) and synapsis (Fig. 1d). In contrast, all these defects were completely rescued in  $(PWD \times B6)F1^{PWD/H}$  hybrids inheriting the engineered humanized zinc-finger array (Fig. 1c, d, Extended Data Fig. 2e). Thus the zinc-finger domain of PRDM9, and hence probably the DNA-binding properties of this protein, underlies the role of *Prdm9* in hybrid sterility.

Although  $(PWD \times B6)F1^{PWD/B6}$  male mice are completely sterile, the male progeny of the reciprocal cross  $(B6 \times PWD)F1^{B6/PWD}$  are semi-fertile<sup>9</sup>. A particular 4.7 Mb locus (*Hstx2*) on the PWD X chromosome influences these fertility differences<sup>6</sup>. We also tested the impact of humanization in this reciprocal cross, and full fertility (from semi-fertility) was again restored (see below and Supplementary Information). Thus humanization of PRDM9 acts at least partially independently of *Hstx2*.

Our reprogramming of the PRDM9 binding sites mimics the consequences of mutational changes in its zinc-finger array. The restoration of hybrid fertility suggests that the same rescue is likely to occur for newly arising alleles that also reset PRDM9 binding sites, and hence hybrid sterility between subspecies driven by *Prdm9* will be evolutionarily transient. This raises the question, which we return to below, of what properties are possessed by *Prdm9* alleles that are associated with reduced fertility.

### Humanizing the recombination landscape

To characterize the consequences of re-engineering the zinc-finger domain on recombination, we generated high-resolution DSB maps for mice with different *Prdm9* alleles and genomic backgrounds, using ChIP-seq single-stranded DNA sequencing<sup>19</sup> on adult testes. This approach identifies single-stranded 3' sequence ends decorated with DMC1, which arise as intermediates following creation of DSBs by SPO11. In addition to mapping DSB hotspots, our hotspot-calling algorithm estimates a hotspot 'heat', proportional to the fraction of cells marked by DMC1 at that locus (Supplementary Information). This DMC1 heat depends on both the relative frequency of DSB formation and on how long DMC1 marks persist<sup>20</sup>. We also obtained complementary information by performing ChIP-seq to measure

H3K4me3, a histone modification directly introduced in *cis* by PRDM9 binding<sup>11</sup>.

Relative to wild-type B6 mice<sup>21</sup>,  $B6^{H/H}$  mice showed completely changed hotspot landscapes (2.6% overlap; Extended Data Fig. 3), with hotspots in the humanized mouse showing strong enrichment for a motif matching the previously reported human PRDM9 binding motif<sup>13</sup> (Extended Data Fig. 4). Most DSB hotspots overlapped H3K4me3 peaks (89%,  $P < 0.05$ ). Correlation between the wild-type and humanized mice in total DSB heats increased over larger genomic scales (Extended Data Fig. 3b), consistent with earlier studies showing large-scale crossover rates depend on factors other than PRDM9<sup>16,20,22</sup>.

In the heterozygous mouse, despite the presence of two different *Prdm9* alleles, we found a similar number of hotspots to homozygous mice (Supplementary Table 1). Furthermore, almost all  $B6^{B6/H}$  hotspots (95.8%) were found in either the  $B6^{B6/B6}$  or  $B6^{H/H}$  mice (Extended Data Figs 3c, 4c, Supplementary Table 2). The human allele exhibited 2.7-fold dominance over the wild-type allele (Supplementary Table 2), with even stronger dominance for hotter hotspots (Extended Data Fig. 3c). Comparison of homozygous and heterozygous hotspot heats (Extended Data Fig. 3d, e) implies B6 hotspots operate similarly, but are proportionally less active, in the heterozygote. For additional DSB hotspot analyses, see Supplementary Information and Extended Data Fig. 5.

### Humanization restores symmetric binding

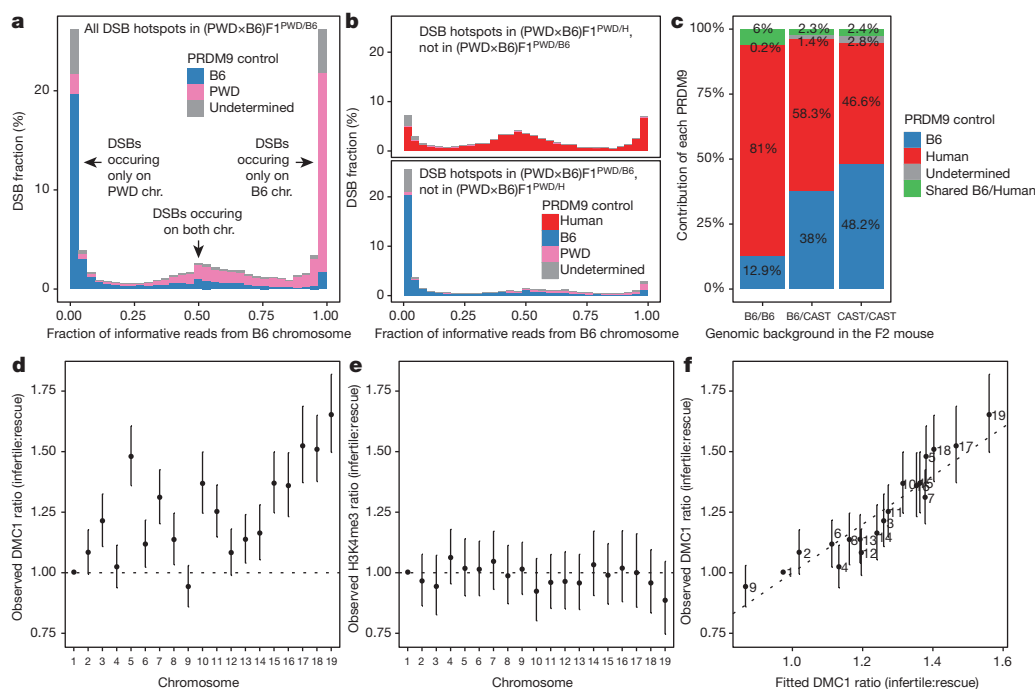
Next we examined DSB hotspot maps for hybrid males: infertile  $(PWD \times B6)F1^{PWD/B6}$ , reciprocal semi-fertile  $(B6 \times PWD)F1^{B6/PWD}$ , humanized rescue  $(PWD \times B6)F1^{PWD/H}$ , and reciprocal humanized rescue  $(B6 \times PWD)F1^{H/PWD}$ , with wild-type PWD for comparison. Sequence differences between the PWD and B6 genomes allowed us to determine whether individual hotspots in these hybrids were 'symmetric', with DSBs occurring equally on both chromosomes, or 'asymmetric', with a preference towards either the PWD or B6 chromosome (Supplementary Information Section 5).

We found that most DMC1 signal (71.8%) in  $(PWD \times B6)F1^{PWD/B6}$  or  $(B6 \times PWD)F1^{B6/PWD}$  hybrids occurs within asymmetric DSB hotspots (Fig. 2a, Extended Data Fig. 6). Further, DSBs associated with the PWD allele occur largely on the B6 chromosome and those associated with the B6 allele occur largely on the PWD chromosome. We also measured asymmetry of the H3K4me3 mark at each hotspot and found the same pattern, confirming that DSB asymmetry largely reflects underlying differences in PRDM9 binding and methylation between the two homologues. This H3K4me3 asymmetry resembles that previously described for  $(B6 \times CAST)F1^{B6/CAST}$  hybrids<sup>17</sup>, but is considerably more extreme. Sequence differences directly disrupting PRDM9 binding motifs explain almost all cases of binding asymmetry (83.4% of PWD hotspots; 91.3% of B6 hotspots), and result from rapid mutational accumulation along the separate lineages from the common ancestor of B6 and PWD (Extended Data Fig. 6g).

Such asymmetry can arise through meiotic drive to favour mutations disrupting PRDM9 binding motifs, within populations where these motifs are active. Any new mutation disrupting PRDM9 binding at a hotspot is preferentially transmitted to offspring: in individuals heterozygous for the mutation, DSBs occur preferentially on the non-mutant chromosome and are then repaired by copying from the mutant chromosome<sup>23</sup>. This phenomenon has been observed at PRDM9 binding motifs in human<sup>13</sup> and mouse<sup>17</sup> and causes a rapid accumulation of mutations disrupting PRDM9 binding. B6 and PWD *Prdm9* alleles are largely subspecies-specific<sup>15</sup>, so only the B6 lineage has experienced strong erosion of the B6 binding motif, and only the PWD lineage has experienced strong erosion of the PWD binding motif. This asymmetric erosion explains the highly asymmetric PRDM9 binding sites in F1 hybrids.

Because the human allele has not been present in mice, its binding sites have not experienced erosion in the mouse genome. As a consequence, DSBs at hotspots attributable to the human allele occur





**Figure 2 | DSB hotspot asymmetry in hybrids.** **a**, Distribution of the fraction of reads originating from the B6 chromosome in the infertile (PWD × B6)F1<sup>PRDM9/B6</sup> mouse. PRDM9 control at each hotspot is attributed to B6 (blue), PWD (pink) or undetermined (grey). **b**, As **a**, but for non-shared hotspots, unique to either the rescue (PWD × B6)F1<sup>PRDM9/H</sup> mouse (top) or the infertile (PWD × B6)F1<sup>PRDM9/B6</sup> mouse (bottom). **c**, Relative contributions of B6 and humanized PRDM9 to DMC1 signal in (B6/CAST)F2<sup>B6/H</sup>. Bars represent the three possible genomic backgrounds.

**d**, Individual chromosome effects (relative to chromosome 1) when comparing DMC1 signals in (PWD × B6)F1<sup>PRDM9/B6</sup> relative to (PWD × B6)F1<sup>PRDM9/H</sup>, for shared DSB hotspots. Error bars, ±1 s.e. **e**, As **d**, but for H3K4me3. **f**, Comparison of DMC1 chromosome effects (as in **d**) with the fitted chromosome effects, using a model including the symmetric hotspot measures for the three *Prdm9* alleles. Error bars, ±3 s.e. (95% simultaneous confidence level for 19 chromosomes).

mostly (57%) in symmetric hotspots, with the remaining, asymmetric hotspots mainly (84.2%) explained by the presence of mutations that coincidentally fall within the human PRDM9 binding motif (Fig. 2b). Conversely, only 30% of DSBs at hotspots attributable to the B6 allele occur in symmetric hotspots. An identical pattern is seen in the reciprocal crosses. Thus, a genome-wide effect of humanizing the mouse is to reprogram hotspot positions with the consequence that hotspot asymmetry is reduced in the hybrids.

Meiotic drive might also explain dominance, as seen for the human *Prdm9* allele over the B6 allele in the B6<sup>B6/H</sup> mouse, because B6 motifs are heavily eroded on the B6 background. To test this, we created F2 mice to analyse the behaviour of the B6 and humanized alleles on a neutral *Mus musculus castaneus* (CAST/EiJ) background which has been unaffected by B6 motif erosion (Extended Data Fig. 6h). Dominance of the human allele disappeared in regions of the genome with two copies of the CAST genome—removing the effect of motif erosion removes the dominance (Fig. 2c). This result excludes some factors which might influence dominance (Supplementary Information), and also suggests that recently arisen *Prdm9* alleles might be dominant over older alleles, for which meiotic drive will have had more time to degrade binding motifs.

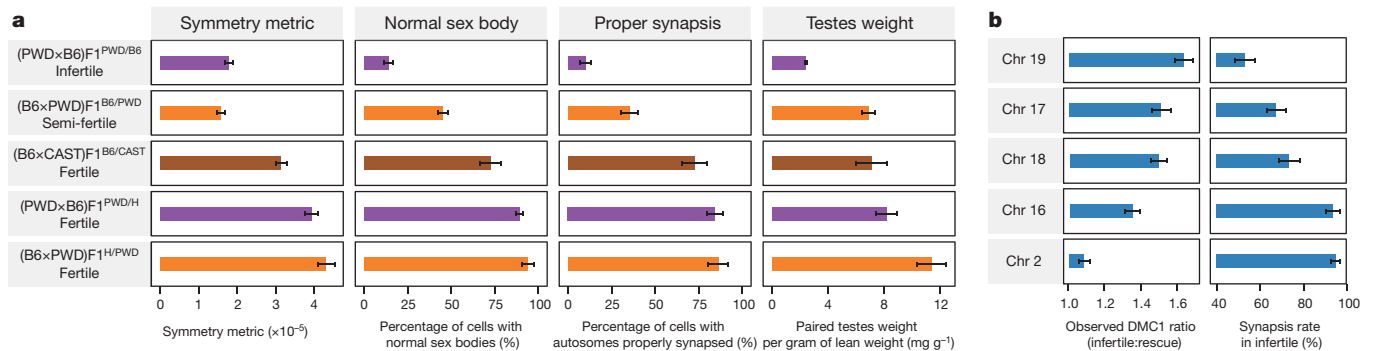
### Chromosome-specific *trans* effects of humanization

The infertile and humanized rescue mice share some hotspots, controlled by the PWD allele. These shared hotspots show strong correlation ( $r^2 = 0.63$ ) in DMC1 heat, but nevertheless far weaker than that between hotspots in the infertile and reciprocal mice (0.95). To explore this weaker correlation, we compared DMC1 heats in the two mice for each shared hotspot and calculated their ratio. We observed substantial differences in these ratios across different chromosomes (Fig. 2d). Thus, substituting the B6 allele for the human allele impacts hotspots that neither allele binds directly, *in trans*, and this impact is observed at broad genomic scales. This *trans* effect might

reflect differences in either the formation, or downstream processing, of DSBs. In contrast to DMC1, the H3K4me3 heat showed no significant chromosomal ratio differences (Fig. 2e), implying that the *trans* effect probably operates downstream of PRDM9 binding. Furthermore, comparison of DMC1 heats between B6<sup>B6/B6</sup> and B6<sup>B6/H</sup> mice also revealed chromosome effects (Extended Data Fig. 7). This implies that such *trans* effects do not depend on SNP presence (the B6 background is fully homozygous), and cannot simply be a consequence of asynapsis (observed only in the infertile mouse).

Next, we sought to understand the drivers of these chromosome-specific differences in DMC1 heat by testing various potential predictors of these differences between the infertile and humanized rescue mice (Supplementary Information). After an exhaustive search over possible models, given the predictors considered, the best-fitting model was highly predictive ( $r^2 = 0.84$ ; Fig. 2f) and included only symmetric hotspot measures—the total H3K4me3 signal from PRDM9 binding on both homologues (that is, symmetrically) at the same hotspots, summed over the entire chromosome—for each of the three *Prdm9* alleles ( $P < 0.01$  in each case). The *trans* effect is thus explained by knowledge of only the direct differences in PRDM9 binding targets across mice, without any additional information regarding other features such as SNP diversity, consistent with the sole difference between the infertile and rescue mice being the zinc-finger array of *Prdm9*. Moreover, only symmetric hotspots (in the infertile mouse, a minority) provide predictive power.

The fitted model implies that lower overall symmetric binding results in increased DMC1 heat, at a chromosome level. The same properties ( $P < 0.0002$ ; Supplementary Information) hold true in the comparison between B6<sup>B6/B6</sup> and B6<sup>B6/H</sup> mice. Although the B6 background is completely homozygous, so PRDM9 is predicted to mark H3K4me3 equally on both homologues, different total levels of H3K4me3 marking across chromosomes still occur and these correlate with observed differences in DMC1 heat between the two genotypes.



**Figure 3 | Humanizing PRDM9 restores proper synapsis and rescues fertility in hybrids.** **a**, Fertility metrics in hybrid mice. Error bars represent bootstrap 95% confidence intervals (symmetry metric), or  $\pm 1$  s.e. (other metrics) **b**, Chromosome effects in DMC1 signals (as Fig. 2d) versus previously reported<sup>5</sup> asynapsis rates for five chromosomes in infertile (PWD  $\times$  B6)F1<sup>PWD/B6</sup>. Error bars,  $\pm 1$  s.e.

This excludes sequence differences at or near hotspots, or asynapsis itself, as a cause, and suggests that the total amount of symmetric binding on each chromosome, as opposed to a simple lack of asymmetric binding, plays an important role in predicting DMC1 heat. The direction of causality is reasonably clear (binding predates DSB formation, and the H3K4me3 mark lacks similar chromosome effects), while confounding influences should always be shared between the mice being compared and thus cannot alone explain the observed inter-chromosomal differences. It therefore appears that differences in the level of overall symmetric binding by PRDM9 drive downstream trans effects at chromosomal scales, with lower symmetric binding somehow increasing the number, or repair time, of DSBs even at distant hotspots.

### PRDM9 binding symmetry and synapsis

Sterile (PWD  $\times$  B6)F1<sup>PWD/B6</sup> hybrids show very high rates of asynapsis, particularly at specific chromosomes<sup>5</sup>, and failure to form the sex body during early meiosis<sup>5,9</sup>. By contrast, these phenotypes are completely rescued in (PWD  $\times$  B6)F1<sup>PWD/H</sup> hybrids harbouring the humanized *Prdm9* allele (Fig. 1c, d). Having seen a relationship between PRDM9 binding symmetry and the recombination process, we examined binding symmetry in relation to fertility. For different male mice, we measured three quantitative fertility phenotypes<sup>24</sup> (Fig. 3a), and calculated several genome-wide measures of hotspot symmetry (Extended Data Fig. 8; Supplementary Information). We observed a significant correlation ( $P = 0.0083$ ; rank correlation permutation test) between the DMC1 symmetry measures and the rate of proper synapsis among all nine mice studied. In humanized hybrid mice, the observed increase in symmetry was accompanied by improved fertility. Notably, this improvement effect is stronger than the *Hstx2* modifier, responsible for the difference in asynapsis and fertility observed between the sterile and reciprocal hybrids<sup>5</sup> (Fig. 3a). An additional mouse hybrid, (B6  $\times$  CAST)F1<sup>B6/CAST</sup>, showed intermediate PRDM9 binding symmetry<sup>17</sup> and also an intermediate asynapsis level. Symmetry measures in homozygous mice (PWD, B6<sup>B6/B6</sup>, B6<sup>B6/H</sup>, B6<sup>H/H</sup>) are, as expected, much higher than hybrids, and these mice show the highest synapsis rates and fertility measures.

Previous work<sup>5</sup> showed that in the infertile (PWD  $\times$  B6)F1 mouse, synapsis failure occurs at different rates among five chromosomes tested. We compared the reported asynapsis rates for these five chromosomes with the chromosome-specific DMC1 heat effects described above and found an identical ranking ( $P = 0.017$  by rank correlation permutation test; Fig. 3b). Because these DMC1 heat effects are strongly predicted by symmetric H3K4me3 levels in the infertile mouse, this result implies that chromosomes with lower symmetric PRDM9 binding experience higher asynapsis rates. This may explain why lower symmetric PRDM9 binding genome-wide accompanies higher overall asynapsis rates among different mice.

Having found elevated DMC1 heat on chromosomes influenced by asynapsis (where homologous pairing fails), we examined DMC1 and

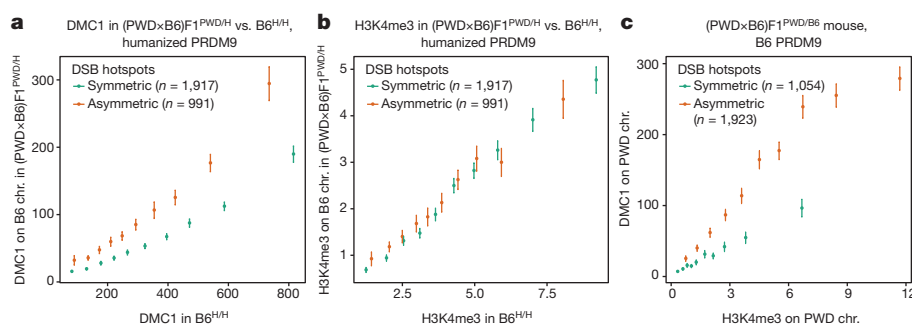
H3K4me3 heats in two additional settings, where no homologue exists at all and thus homologous chromosome pairing cannot occur: the X chromosome in male mice, and separately in humanized hybrid mice at autosomal hotspots where the human PRDM9 binding motif lies within a region deleted in the PWD genome. In both these settings, we observed an elevation of DMC1 heat relative to autosomal hotspots bound symmetrically by PRDM9 (Extended Data Fig. 9). Elevation of DMC1 heat might, therefore, be a consistent signature of non-pairing of homologous chromosomes during meiosis. DMC1 elevation might be explained by an increased probability of a DSB occurring at that site, or by the DMC1 coating at breaks persisting for longer (delayed repair). However, the total number of RAD51-marked DSBs initiated per cell is tightly regulated<sup>25</sup>, remaining unchanged even in *Prdm9* knockouts<sup>26</sup>, while in both knockouts and infertile hybrids, DSB marks indeed persist late into pachytene, suggesting a failure of repair<sup>5,9,26</sup>. Therefore, the elevated DMC1 signals we observe may be explained by persistence of DMC1 where homologous repair is compromised or delayed.

### PRDM9-dependent homologue interactions

Given our chromosome-scale observations, we next asked whether symmetric binding at individual hotspots might also influence DMC1 heat. At each human-controlled hotspot in the humanized rescue, we measured the component of total DMC1 heat contributed by the B6 chromosome only, and compared this to the DMC1 heat for the same hotspot in B6<sup>H/H</sup>. The comparison revealed (Fig. 4a) a remarkably strong, and clear, elevation in DMC1 heat in the hybrid mouse for the asymmetric hotspots ( $>90\%$  asymmetry, towards binding of only the B6 chromosome), relative to the symmetric hotspots (those within 10% of complete symmetry). However, similar to the chromosomal analysis, H3K4me3 enrichment showed no difference whatsoever between symmetric and asymmetric sites in these mice (Fig. 4b). Indeed a comparison of H3K4me3 and DMC1 heat revealed a far higher (Fig. 4c) ratio of average DMC1 heat to H3K4me3 enrichment for asymmetric relative to symmetric hotspots, across all hybrid mice, backgrounds, and *Prdm9* alleles tested (Extended Data Fig. 9d). This effect reflects a consistent elevation of DMC1 heat at DSB sites on individual chromosomes when the homologue is not bound strongly (Extended Data Fig. 9e, f). This phenomenon cannot easily be explained by factors including local heterozygosity within or outside the PRDM9 motif, the type of mutation(s) disrupting PRDM9 binding, or outlier effects (Extended Data Figs 9, 10; Supplementary Information Section 13).

Thus, elevation of DMC1 heat on the bound chromosome appears to be a universal feature of hotspots where PRDM9 binds asymmetrically, relative to symmetrically bound hotspots. By contrast, the results for H3K4me3 suggest the mark is deposited in an independent manner on each homologue (Supplementary Information Section 12.2). This implies the DMC1 heat elevation depends on a process involving symmetric PRDM9 binding, downstream of H3K4me3 deposition, involving both homologues. While we cannot exclude the possibility





**Figure 4 | Asymmetric DSB hotspots show elevated DMC1 signals but not H3K4me3 elevation.** **a**, Comparison of B6<sup>H/H</sup> and (PWD × B6)F1<sup>PWD/H</sup> DMC1 signals (medians shown). Signals are compared for symmetric and asymmetric hotspots in (PWD × B6)F1<sup>PWD/H</sup>, on the shared B6 chromosomes. Error bars, 95% confidence intervals. **b**, As **a**, but

for H3K4me3. **c**, Comparison of DMC1 signals in (PWD × B6)F1<sup>PWD/B6</sup>, at symmetric and asymmetric hotspots, binned by H3K4me3 enrichment (medians shown). H3K4me3 and DMC1 signals are estimated on the PWD chromosome only, for hotspots associated with B6 PRDM9.

that somehow more DSBs occur at asymmetric hotspots, this would require early, precise pairing of homologues, at least at hotspots, before DSB formation, to determine which hotspots are symmetrically bound. Although there is some evidence of pre-meiotic homologue association<sup>27</sup>, current data do not suggest the existence of precise pairing before DSB formation<sup>28</sup>. The alternative and more plausible explanation is that sites where PRDM9 binds asymmetrically simply experience a delay in DSB processing, delaying DMC1 removal compared to symmetric DSB hotspots. While our data represent the collective behaviour of populations of cells, this model suggests a mechanism of PRDM9-dependent interaction between homologues influencing downstream DSB processing operating within individual cells, which we discuss below (also Supplementary Information Section 14).

## Discussion

Only one mammalian speciation gene, *Prdm9*, has so far been identified. Humanizing the zinc-finger array of *Prdm9* redirects binding, thereby entirely reprogramming recombination hotspots, and in doing so reverses the hybrid infertility between *musculus* and *domesticus* subspecies. This modification mimics the consequences of a newly arising allele and thus suggests that *Prdm9* evolution (for example, rapid fixation of particular existing variants<sup>3,15</sup> or novel alleles arising by mutation) in either or both subspecies would also restore hybrid fertility.

Multiple lines of evidence in our data, at chromosomal, whole-organism, and individual hotspot scales, strongly suggest novel roles for PRDM9 in the formation or processing of DSBs downstream of H3K4me3 deposition, dependent upon symmetric binding. Several aspects of our, and published, data (comparison between B6<sup>B6/B6</sup> and B6<sup>B6/H</sup> mice, see also Supplementary Information) also mean that our results cannot be fully explained simply by sequence differences within or around hotspots, which do not specifically impact binding symmetry.

Pervasive asynapsis is proposed to be the underlying cause of infertility in hybrid mice<sup>5</sup>. We observed a positive relationship between symmetric PRDM9 binding and correct synapsis of homologous chromosomes later in meiosis. Replacing the B6 allele with the humanized allele in hybrids greatly increases symmetric binding, restoring proper synapsis and fertility. Many apparently complex relationships have previously been reported between naturally occurring mouse *Prdm9* alleles, allelic dosage, and quantitative fertility measures in hybrids<sup>9</sup>. Each of ten manipulations shown or predicted to increase PRDM9 binding symmetry also increases meiotic success and fertility (Supplementary Information), supporting the idea that the link between binding symmetry and fertility might be very general, and causal.

The erosion of PRDM9 binding sites through meiotic drive<sup>17</sup> also occurs at human hotspots<sup>13</sup>, and probably across many mammals. In two populations separated for sufficient time, differential PRDM9 binding site erosion will decrease symmetry in hybrids, which is likely

to decrease fertility levels (though not necessarily to the extreme of sterility). Therefore, PRDM9 may affect hybrid fertility levels across many mammalian species and so might repeatedly act in driving early speciation steps, although the rapid evolution of the zinc-finger array of PRDM9 implies an unexpected transience of this direct role. However, even subtle or transient PRDM9-driven reductions in fertility might still provide a selective advantage to additional mutations contributing towards speciation. This mechanism is different from the previously characterized causes of intrinsic hybrid incompatibilities, such as differences in ploidy, chromosomal rearrangements, or incompatibilities between genes. The extent to which it has been responsible for speciation in the natural world appears an interesting question for further research.

One plausible mechanism for the impacts of (a)symmetry involves a role for PRDM9 binding in aiding homology search—a process thought to involve invasion of the homologous chromosome to probe for homology by single-stranded DNA formed around DSBs<sup>29</sup>. It has been suggested that synaptonemal complex proteins are loaded at some DSB sites and synapsis begins to spread<sup>7,8</sup>. Extending this model, to incorporate the property that asymmetrically bound sites are less favourable for homology search, would parsimoniously predict each symmetry-related phenomenon we observed: DSBs at asymmetric hotspots would repair more slowly, elevating their DMC1 signal, and chromosomes with fewer symmetric hotspots overall would show delayed DSB repair and higher asynapsis rates, ultimately causing subfertility or sterility in animals with low symmetric binding. It is not known how homology search occurs efficiently in the nuclear environment, given the enormous potential search space of the genome<sup>30</sup>, or why hotspots exist at all. Both phenomena could be explained by the above model in which homology search is focused at least partly on hotspot positions. Indeed hotspots might massively increase search efficiency by directing homology search to PRDM9 binding sites.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 16 July; accepted 21 December 2015.**

**Published online 3 February 2016.**

- Presgraves, D. C. The molecular evolutionary basis of species formation. *Nature Rev. Genet.* **11**, 175–180 (2010).
- Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J. C. & Forejt, J. A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* **323**, 373–375 (2009).
- Oliver, P. L. et al. Accelerated evolution of the *Prdm9* speciation gene across diverse metazoan taxa. *PLoS Genet.* **5**, e1000753 (2009).
- Gregorová, S. & Forejt, J. PWD/Ph and PWK/Ph inbred mouse strains of *Mus m. musculus* subspecies—a valuable resource of phenotypic variations and genomic polymorphisms. *Folia Biol. (Praha)* **46**, 31–41 (2000).
- Bhattacharyya, T. et al. Mechanistic basis of infertility of mouse intersubspecific hybrids. *Proc. Natl Acad. Sci. USA* **110**, E468–E477 (2013).

6. Bhattacharyya, T. *et al.* X chromosome control of meiotic chromosome synapsis in mouse inter-subspecific hybrids. *PLoS Genet.* **10**, e1004088 (2014).
7. Qiao, H. *et al.* Interplay between synaptonemal complex, homologous recombination, and centromeres during mammalian meiosis. *PLoS Genet.* **8**, e1002790 (2012).
8. Henderson, K. A. & Keeney, S. Synaptonemal complex formation: where does it start? *Bioessays* **27**, 995–998 (2005).
9. Flachs, P. *et al.* Interallelic and intergenic incompatibilities of the *Prdm9* (*Hst1*) gene in mouse hybrid sterility. *PLoS Genet.* **8**, e1003044 (2012).
10. Hayashi, K., Yoshida, K. & Matsui, Y. A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. *Nature* **438**, 374–378 (2005).
11. Grey, C. *et al.* Mouse PRDM9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination. *PLoS Biol.* **9**, e1001176 (2011).
12. Jeffreys, A. J., Kauppi, L. & Neumann, R. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nature Genet.* **29**, 217–222 (2001).
13. Myers, S. *et al.* Drive against hotspot motifs in primates implicates the *PRDM9* gene in meiotic recombination. *Science* **327**, 876–879 (2010).
14. Berg, I. L. *et al.* *PRDM9* variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nature Genet.* **42**, 859–863 (2010).
15. Kono, H. *et al.* *Prdm9* polymorphism unveils mouse evolutionary tracks. *DNA Res.* **21**, 315–326 (2014).
16. Auton, A. *et al.* A fine-scale chimpanzee genetic map from population sequencing. *Science* **336**, 193–198 (2012).
17. Baker, C. L. *et al.* *PRDM9* drives evolutionary erosion of hotspots in *Mus musculus* through haplotype-specific initiation of meiotic recombination. *PLoS Genet.* **11**, e1004916 (2015).
18. Forejt, J. & Ivanyi, P. Genetic studies on male sterility of hybrids between laboratory and wild mice (*Mus musculus* L.). *Genet. Res.* **24**, 189–206 (1974).
19. Khil, P. P., Smagulova, F., Brick, K. M., Camerini-Otero, R. D. & Petukhova, G. V. Sensitive mapping of recombination hotspots using sequencing-based detection of ssDNA. *Genome Res.* **22**, 957–965 (2012).
20. Pratto, F. *et al.* DNA recombination. Recombination initiation maps of individual human genomes. *Science* **346**, 1256442 (2014).
21. Brick, K., Smagulova, F., Khil, P., Camerini-Otero, R. D. & Petukhova, G. V. Genetic recombination is directed away from functional genomic elements in mice. *Nature* **485**, 642–645 (2012).
22. Paigen, K. *et al.* The recombinational anatomy of a mouse chromosome. *PLoS Genet.* **4**, e1000119 (2008).
23. Jeffreys, A. J. & Neumann, R. Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. *Nature Genet.* **31**, 267–271 (2002).
24. Flachs, P. *et al.* *Prdm9* incompatibility controls oligospermia and delayed fertility but no selfish transmission in mouse intersubspecific hybrids. *PLoS ONE* **9**, e95806 (2014).
25. Lange, J. *et al.* ATM controls meiotic double-strand-break formation. *Nature* **479**, 237–240 (2011).
26. Sun, F. *et al.* Nuclear localization of PRDM9 and its role in meiotic chromatin modifications and homologous synapsis. *Chromosoma* (2015).
27. Boateng, K. A., Bellani, M. A., Gregoretti, I. V., Pratto, F. & Camerini-Otero, R. D. Homologous pairing preceding SPO11-mediated double-strand breaks in mice. *Dev. Cell* **24**, 196–205 (2013).
28. Ishiguro, K. *et al.* Meiosis-specific cohesin mediates homolog recognition in mouse spermatocytes. *Genes Dev.* **28**, 594–607 (2014).
29. Renkawitz, J., Lademann, C. A. & Jentsch, S. Mechanisms and principles of homology search during recombination. *Nature Rev. Mol. Cell Biol.* **15**, 369–383 (2014).
30. Weiner, A., Zauberman, N. & Minsky, A. Recombinational DNA repair in a cellular context: a search for the homology search. *Nature Rev. Microbiol.* **7**, 748–755 (2009).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank N. Hortin, S. Chen and R. Davies for technical assistance, the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics for the generation of the sequencing data and R. Esnouf and J. Diprose for assistance with computing facilities. PWD/PhJ mice were a gift from J. Forejt. This work was supported by the Wellcome Trust Core Award Grant 090532/Z/09/Z, Senior Investigator Award 095552/Z/11/Z (to P.D.), Investigator Award 098387/Z/12/Z (to S.R.M.) and the NIDDK Intramural Research Program (R.D.C.O.). E.H. is funded by a Nuffield Department of Medicine Prize Studentship. J.G.H. is an EPAC/Linacre Junior Research Fellow funded by the Human Frontiers Postdoctoral Program (LT-001017/2013-L).

**Author Contributions** S.R.M. and P.D. designed the study. B.D., N.A., E.B., D.B., R.D. and C.P. generated and bred the transgenic mice. F.P., G.Z. and R.D.C.O. performed and oversaw DMC1 ChIP-seq. N.A. performed H3K4me3 ChIP-seq. D.M. and C.G. performed cytological analysis. E.H., J.G.H., N.A., A.G.H., R.L. and K.B. analysed the data. B.D., E.H., J.G.H., N.A., A.G.H., S.R.M. and P.D. wrote the paper, with input from all authors.

**Additional Information** The source data generated in this publication has been deposited in NCBI's Gene Expression Omnibus under accession number GSE73833. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.R.M. ([myers@stats.ox.ac.uk](mailto:myers@stats.ox.ac.uk)) or P.D. ([donnelly@well.ox.ac.uk](mailto:donnelly@well.ox.ac.uk)).



## METHODS

**Gene targeting in embryonic stem cells.** A C57BL/6J (B6) mouse genomic BAC clone (RP23-159N6) encompassing the *Prdm9* gene was used for subcloning of homology regions. A 7 kb *XmaI*/*SpeI* fragment upstream of exon 10 and a 2.5 kb *BamHI*/*SpeI* fragment downstream of exon 10 were used as 5' and 3' homology regions, respectively. The intervening 4 kb *SpeI*/*BamHI* encoding exon 10 and flanking intronic regions were subcloned and an internal 1.4 kb *BglII*-*NheI* fragment, containing the coding region of the zinc-finger array, was replaced with a synthesized fragment (Life Technologies) encoding the zinc-finger array from the human B allele. All coding sequence 5' of the first zinc finger and all 3' untranslated regions (UTR) downstream of the stop codon were left as mouse. This humanized fragment was then assembled between the two homology arms, upstream of a neomycin selection cassette. *PhlC31* attP sites were incorporated immediately downstream of the 5' homology arm and between the PGK promoter and the neomycin phosphotransferase open reading frame to equip the locus with *PhlC31* integrase cassette exchange machinery for subsequent manipulations<sup>31</sup>.

The completed targeting vector was linearized with *Apal* and electroporated into mycoplasma free C57BL/6N JM8F6 embryonic stem cells (Extended Data Fig. 1a). JM8F6 cells were a gift from B. Skarnes, Wellcome Trust Sanger Institute. Following selection in  $210 \mu\text{g ml}^{-1}$  G418, recombinant clones were screened by PCR to detect homologous recombination over the 3' arm. A forward primer (5'-TACCGGTGGATGTGGAATGTG-3') binding within the PGK promoter was used together with a reverse primer (5'-TGACAGCAAAAACCACTCTA-3') binding downstream of the 3' homology arm to amplify a 2.7 kb fragment from correctly recombined clones. Positive clones were examined for correct recombination at the 5' end by long range PCR using a forward primer (5'-CAGAGGACCTTTAGTCTGTGAGGG-3') binding upstream of the 5' homology arm and a reverse primer (5'-AGCAGAGGCTTGACCTATCGCTAA-3') binding within the humanized region. Correctly targeted clones yielded a 10.4 kb amplicon. Sanger sequence analysis of the 10.4 kb amplicon encompassing the 5' homology arm with primer 5'-CCTTTCTCAATGATCCACAAAT-3' confirmed the correct integration of the 5' attP sequence, necessary for future manipulations of the locus. Southern blotting using a probe against neomycin was used to confirm that only a single integration event had occurred.

**Mouse production and matings.** Mice were housed in individually ventilated cages and received food and water *ad libitum*. All studies received local ethical review approval and were performed in accordance with the UK Home Office Animals (Scientific Procedures) Act 1986. Experimental groups were determined by genotype and were therefore not randomized, with no animals excluded from the analysis. Sample size for fertility studies and cytogenetics (see below) were selected on the basis of previously published studies<sup>5,9,32</sup>. No statistical methods were used to predetermine sample size. All phenotypic characterization was performed blind to experimental group.

ES cells from correctly targeted clones were injected into albino C57BL/6J blastocysts and the resulting chimaeras were mated with albino C57BL/6J females. Successful germline transmission yielded black pups and F1 mice harbouring the humanized *Prdm9* allele were identified using the above attP screening PCR. F1 heterozygous male mice were bred with C57BL/6J *Flp* recombinase deleter mice (*Tg*(*ACTB-Flpe*)9205Dym (Jax stock 005703)) and offspring were screened for the deletion of the selection cassette using a forward primer (5'-TTCTGCCATCACTTCCTTCGGTGA-3') binding immediately upstream of the cassette and a reverse primer (5'-TCTGAAGCCCAACTTTCATTAATACCCC-3') binding immediately downstream of the cassette. A 677-bp amplicon was obtained from the *Flp*-deleted humanized allele and a 491-bp amplicon was obtained from the wild-type allele. Heterozygous humanized mice without the selection cassette were then backcrossed with C57BL/6J to remove the *Flp* transgene before intercrossing to obtain experimental cohorts of heterozygous, homozygous and wild-type mice which were genotyped with the above PCR. PWD/PhJ mice were a gift from J. Forejt, Institute of Molecular Genetics, Prague, Czech Republic and CAST/EiJ were sourced from MRC Harwell.

Fertility was assessed in male mice between the ages of 2 and 4 months by recording the average number of pups obtained when bred with 7-week-old wild-type C57BL/6J female mice. Paired testes weight was recorded and normalized against lean body weight, as assessed using EchoMRI-100 Small Animal Body Composition Analyzer.

**Immunohistochemistry analyses.** Spermatocytes from mice at approximately 9 weeks of age were prepared for immunohistochemistry by surface spreading<sup>33,34</sup>. In brief, the testis tunica was removed, the tubules cut with a razor blade and disassembled by pipetting, in PBS, containing protease inhibitors (Complete, Roche). Following centrifugation at 5,800 g for 5 min, the cells were resuspended in 0.1 M sucrose, and spread onto the surface of slides in a drop of 1% paraformaldehyde

in PBS. The slides were left to dry for 3 h at room temperature, in a humidified box, then washed in 0.4% Photo-Flo 200 (Kodak), and either used immediately, or stored at  $-80^{\circ}\text{C}$ . For immunohistochemistry the following antibodies were used: mouse anti-MLH1 (BD, 51-1327GR); mouse anti-phospho-H2A.X (Millipore 05-636, clone JBW301); rabbit anti-SYCP1 (Novus Biological, NB300-229); rabbit anti-DMC1 (Santa Cruz Biotechnology sc-22768, H-100); mouse anti-SYCP3 (Santa Cruz Biotechnology sc-74569, D-1); rabbit anti-SYCP3 (Abcam ab15093). Non-specific binding sites were blocked by incubating the cells with 0.2% BSA, 0.2% gelatin, 0.05% Tween-20 in PBS (B/ABD buffer). Cells were incubated with the primary antibodies overnight at  $4^{\circ}\text{C}$ . Following washes in B/ABD buffer and detection with secondary antibodies, the slides were mounted in DAPI/Vectashield (Oncor) and analysed with an Olympus BX60 microscope for epifluorescence, equipped with a Sensys CCD camera (Photometrics, USA), using Genus Cytovision software (Leica).

Spermatocytes were staged based on SYCP3 staining. For MLH1 analysis, only pachytenes with 19 or more foci, co-localizing with SYCP3, were considered, according to criteria defined by ref. 35. For DMC1 analysis, randomly selected cells, from any stage, were scored. The number of DMC1 foci per cell was counted using the PointPicker macro in ImageJ (64 bit). For SYCP1 analysis, only cells in pachytene were considered. Cells with 19 fully synapsed autosomes, with co-localizing SYCP1 and SYCP3 signals, and one XY body, were considered normal. For characterization of  $\gamma\text{H2AX}$ , cells in pachytene or diplotene were scored, and we considered normal those where only a clearly identifiable XY body was covered by  $\gamma\text{H2AX}$  signal.

***Prdm9* expression via RT-PCR analysis.** To verify the correct expression of the humanized *Prdm9*, we performed exon-spanning end-point RT-PCR on whole testis cDNA prepared using Tetro reverse transcriptase (Bioline) using a forward primer binding to exon 9 (5'-CATTAAG TGGGGAAGCAAGA-3') and a reverse primer binding within the 3' UTR, immediately downstream of the humanized zinc-finger domain encoded by exon 10 (5'-GGGATTAAATTCCTTTTCTAGTCA-3') (Extended Data Fig. 1b). Q-PCR analysis of *Prdm9* transcripts was performed using two primer pairs (5'-GAATGAGAAAGCAACAGCA-3' and 5'-GGACAACCAAGACTGCACAGA-3'; 5'-AGCAACAGCAATAAAACCA-3' and 5'-GGGATTAAATTCCTTTTCTAGTCA-3'), amplifying regions within the 3' UTR, normalizing against a housekeeping gene (*Hprt*; 5'-AGCTACTGTAATGATCAGTCAACG-3' and 5'-AGAGGTCTCTTTT CACCAGCA-3') using the Power SYBR Green PCR Master mix (Applied Biosystems) and a BioRad CFX96 cyclor as per the manufacturer's instructions. Relative expression was calculated using the Livak method. Expression of the humanized *Prdm9* allele was unaffected by the genetic manipulation (Extended Data Fig. 1b, c).

**Single-stranded DNA sequencing and double-strand break (DSB) detection.** Testis cells from  $\text{B6}^{\text{H/H}}$ ,  $\text{B6}^{\text{B6/H}}$ , wild-type PWD, the infertile ( $\text{PWD} \times \text{B6}$ )  $\text{F1}^{\text{PWD/B6}}$ , the reciprocal semi-fertile ( $\text{B6} \times \text{PWD}$ )  $\text{F1}^{\text{PWD/B6}}$ , the humanized rescue ( $\text{PWD} \times \text{B6}$ )  $\text{F1}^{\text{PWD/H}}$ , ( $\text{B6} \times \text{CAST}$ )  $\text{F1}^{\text{B6/CAST}}$ , ( $\text{B6/CAST}$ )  $\text{F2}^{\text{B6/H}}$  males were subjected to single-stranded DNA sequencing (SSDS) as previously described<sup>19</sup>. In addition, we used the sample C57BL/6 (sample 1) from ref. 21 aligned to mm9/NCBI37. This sample was also re-mapped to mm10/NCBI38 with a modified BWA mapper<sup>19</sup>. Other samples from ref. 21, 9R (sample 2), 13R (samples 1 and 2) and *Prdm9* knockout ( $\text{B6}^{-/-}$ ) (sample 1)<sup>10</sup>, were also used in the comparative analysis of DSB maps (Extended Data Fig. 3e).  $\text{B6}^{\text{H/H}}$  and  $\text{B6}^{\text{B6/H}}$  libraries were prepared in D. Camerini-Otero's lab (NIH) and sequenced on a HiSeq 2000 platform, using paired-end reads (read 1: 36 bp; read 2: 40 bp). These samples were aligned to the mouse mm9/NCBI37 reference genome.

Wild-type PWD, the infertile ( $\text{PWD} \times \text{B6}$ )  $\text{F1}^{\text{PWD/B6}}$ , the reciprocal semi-fertile ( $\text{B6} \times \text{PWD}$ )  $\text{F1}^{\text{PWD/B6}}$ , the reciprocal rescue ( $\text{PWD} \times \text{B6}$ )  $\text{F1}^{\text{PWD/H}}$ , ( $\text{B6} \times \text{CAST}$ )  $\text{F1}^{\text{B6/CAST}}$ , ( $\text{B6/CAST}$ )  $\text{F2}^{\text{B6/H}}$  samples were prepared in The Wellcome Trust Centre for Human Genetics and sequenced on HiSeq 2000 and HiSeq 2500 platforms, using paired-end reads (51 bp for each read). These samples were aligned to the mouse mm10/NCBI38 reference genome with a modified BWA mapper<sup>19</sup>. Variation in the number of sequenced fragments results from the difficulty to precisely assess the DNA concentration before sequencing. Only fragments with high mapping quality (at least 20) were retained for DSB hotspot calling, and only one copy of each duplicate fragment was conserved (here, a fragment is duplicated if there exists at least one other fragment mapping to the same genomic position). Supplementary Table 1 gives details about the samples considered in this study.

**H3K4me3 ChIP-seq.** ChIP-seq was performed as previously described<sup>36</sup> with several modifications (noted here). In brief, the testis tunica was removed, the tubules disassociated with tweezers and fixed in 1% formaldehyde in PBS for 5 min followed by glycine quenching (125 mM final conc.) for 5 min at room temperature. Following washing steps, pellets were resuspended in 900  $\mu\text{l}$  cold RIPA lysis

buffer, dounced 20 times and sonicated in 300 µl aliquots in a Bioruptor Twin sonication bath at 4 °C for three 10-min periods of 30 s on, 30 s off at high power, then cell debris was pelleted and removed and aliquots were pooled. For each sample, 50 µl of equilibrated magnetic beads were resuspended in 100 µl PBS/BSA and added to the chromatin samples for pre-clearing for 2 h at 4 °C with rotation. Beads were removed, and 100 µl of pre-cleared chromatin was set aside for the input control. 5 µl rabbit polyclonal anti-H3K4me3 antibody (Abcam ab8580) was added to the remaining pre-cleared chromatin and incubated overnight at 4 °C with rotation. 50 µl beads were washed and resuspended as before, then incubated with the chromatin samples for 2 h at 4 °C with rotation. Beads were then washed and de-crosslinked at 65 °C as described<sup>36</sup>, and for input controls, 50 µl of pre-cleared chromatin was used. After de-crosslinking, samples were further incubated with 80 µg RNase A at 37 °C for 60 min and then with 80 µg Proteinase K at 55 °C for 90 min. DNA was purified using a Qiagen MinElute reaction cleanup kit.

ChIP and total chromatin DNA samples were sequenced in multiplexed paired-end Illumina libraries, yielding 51-bp reads. We prepared two biological replicates plus one genomic input control each for the infertile (PWD × B6)F1<sup>PWD/B6</sup>, reciprocal (B6 × PWD)F1<sup>B6/PWD</sup>, and rescue (PWD × B6)F1<sup>PWD/H</sup> mice, yielding roughly 40–50 million usable read pairs per replicate. For the B6<sup>B6/B6</sup> and B6<sup>H/H</sup> mice, we prepared one biological replicate each (yielding 70–80 million usable read pairs per sample) and later split read pairs into pseudoreplicates. Sequencing reads were aligned to mm10 using BWA aln<sup>37</sup> (v. 0.7.0) followed by Stampy<sup>38</sup> (v. 1.0.23, option bamkeepgoodreads), and reads not mapped in a proper pair with insert size smaller than 10 kb were removed. Read pairs representing likely PCR duplicates were also removed by samtools rmdup. Pairs for which neither read had a mapping quality score greater than 0 were removed. Fragment coverage was computed at each position in the genome and in 100-bp non-overlapping bins using in-house code and the samtools<sup>39</sup> and bedtools<sup>40</sup> packages.

**DSB hotspot detection and map comparison.** To analyse DMC1 data, we developed a novel ChIP-seq peak caller, specific to DSB hotspots, which takes advantage of the shift in the mapping of single stranded DNA (ssDNA) reads between the 5' and the 3' DNA strands to call hotspots. These ssDNA segments are a consequence of the resection of DNA ends that accompanies a DSB and are isolated by DMC1 ChIP<sup>19</sup>. For each hotspot, the caller estimates in particular the centre of the hotspot, and its heat, loosely defined as the number of reads mapping to this DSB hotspot and predicted to represent real signal. The caller handles sample replicates and is able to call hotspots using several samples jointly. Details are given in Supplementary Information. DSB hotspots from two different samples are considered to overlap if their centres are at most 600 bp apart. DMC1 hotspot heats have been normalized so that the sum of hotspot heats is identical in each sample (and equals the sum of hotspot heats in B6<sup>B6/B6</sup> (sample 1)).

H3K4me3 enrichments have been computed at DSB hotspots identified by DMC1 ChIP-seq, using our previously published method<sup>36</sup> (Supplementary Information Subsection 7.1). H3K4me3 hotspots have also been called *de novo*, without using DSB hotspots, using the same approach<sup>36</sup>. The *de novo* calls were used to generate a list of regions likely to be trimethylated independently of PRDM9, by intersecting calls in mice with different *Prdm9* alleles. In comparisons involving both DMC1 and H3K4me3 data, we excluded DSB hotspots contained in any of the PRDM9-independent trimethylated regions, and we used H3K4me3 enrichments computed at DSB hotspots (Supplementary Information). We only used *de novo* calls for analysis in Extended Data Fig. 6d, e.

**DNA binding motif analyses.** We developed a new, Bayesian, approach to identify DNA motifs enriched at DSB hotspots (Supplementary Information). We used FIMO (MEME Suite version 4.9.1) to find the locations of those motifs genome-wide. Using *Mus famulus* and *Mus caroli* as outgroups, we reconstructed an ancestral reference genome for B6 and PWD. We could therefore identify on which lineage (B6 or PWD) mutations between these two mouse strains occurred. See Supplementary Information for details.

**DSB hotspot assignment in hybrids.** Using SNPs between the B6 and PWD genomes, each read pair from a hybrid DSB library (DMC1 ChIP-seq) is assigned to one of the categories 'B6', 'PWD', 'unclassified' or 'uninformative' using criteria detailed in Supplementary Information. For each DSB hotspot, the ratio of

informative reads from the B6 chromosome was then computed as the fraction of 'B6' reads mapped within 1 kb of the hotspot centre, over the sum of 'B6' and 'PWD' reads in that region. We followed a similar approach for H3K4me3 ChIP-seq, but we further corrected for background signal.

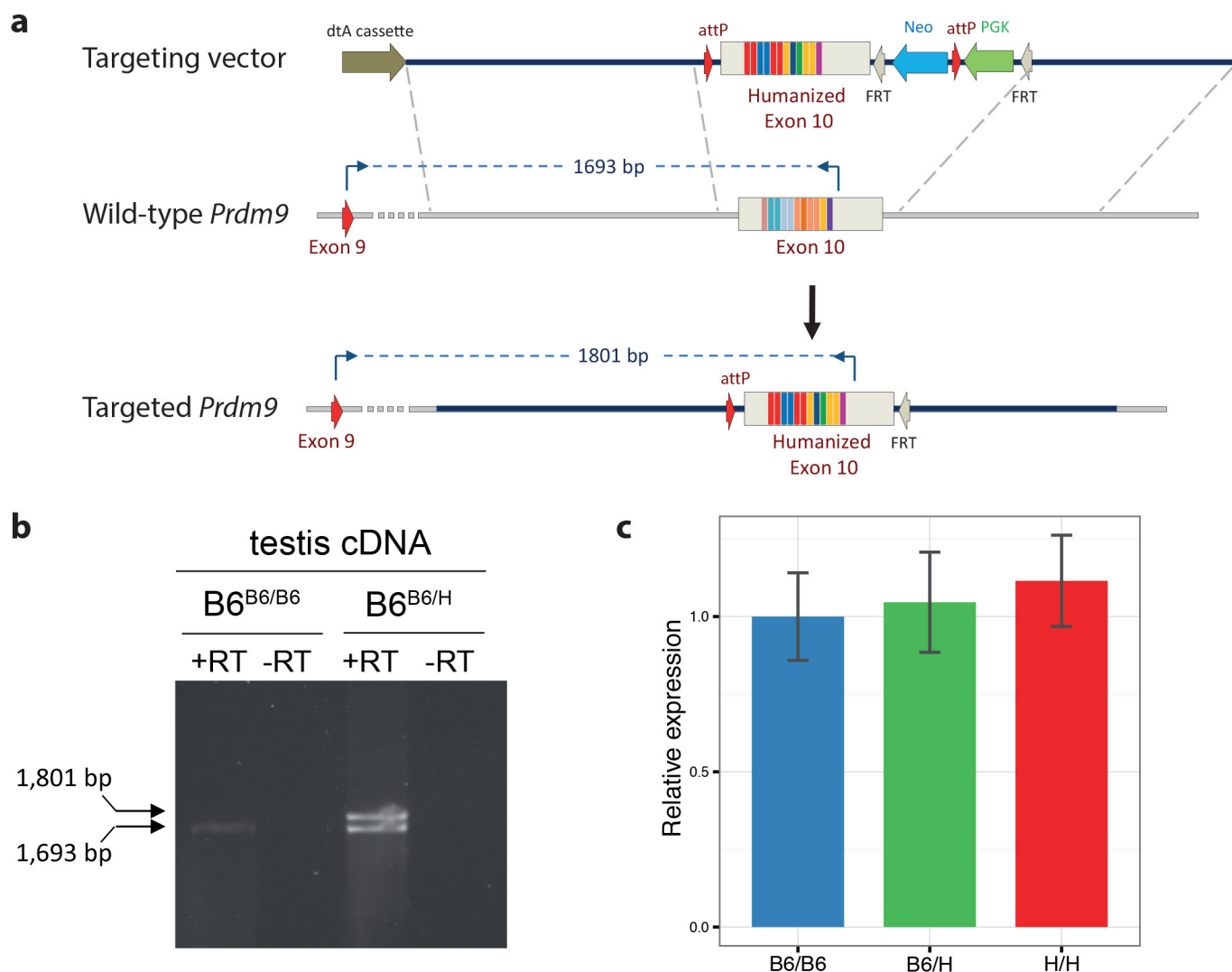
**Chromosome effects.** To test for statistically significant differential elevation of DMC1 (or H3K4me3) heats between chromosomes following *Prdm9* humanization of the infertile (PWD × B6)F1<sup>PWD/B6</sup> mice, we fitted a quasi-Poisson model to these heats, including predictors for each chromosome. Specifically, we estimated parameters  $\alpha$ ,  $\beta$  and  $\gamma$  by fitting the model  $\log[\mathbb{E}(d_{\text{infertile}}|d_{\text{rescue}}, c)] = \alpha + \gamma \log[d_{\text{rescue}}] + \sum_{i=1}^{19} \beta_i^c 1_{\{c=i\}}$ , where  $d_{\text{infertile}}$  and  $d_{\text{rescue}}$  are the DMC1 heats of a particular hotspot which is shared between the infertile (PWD × B6)F1<sup>PWD/B6</sup> and rescue (PWD × B6)F1<sup>PWD/H</sup> mice and  $c$  is a categorical variable which represents the chromosome on which the DSB hotspot occurs. Furthermore, for one of the hybrid mice we considered, for a given autosome, we defined the 'total H3K4me3 signal from PRDM9 binding on both homologues (that is, symmetrically) at the same hotspots, summed over the entire chromosome', also referred to as 'the sum of symmetric' heats', as  $\sum_i 4r_i(1 - r_i)h_i^2$ , where  $r_i$  is the fraction of DMC1 reads coming from the B6 chromosome for hotspot  $i$ ,  $h_i$  is the H3K4me3 heat of that hotspot, and the sum is taken over all the hotspots on that chromosome which are under the control of a specific (PWD, B6, or humanized) PRDM9. (Our analyses always refer to this sum of symmetric heats for a specific allele.) When we considered the B6 mouse (which of course has two B6 chromosomes), we defined this sum of symmetric heats to be  $\sum_i h_i^2$  (which is the special case of the formula above with  $r_i = \frac{1}{2}$ , corresponding to all hotspots being fully symmetric). Under the assumptions we describe in the Supplementary Information, this can also be interpreted as being proportional to the expected number of hotspots with PRDM9 bound on both homologues. Details and motivations for defining this quantity are given in Supplementary Information Section 8, together with a slight adjustment we used in practice to provide robustness against outliers in the value of  $h_i^2$ .

We proceeded similarly in the B6<sup>B6/B6</sup>–B6<sup>B6/H</sup> comparison. The observed effects reported in Fig. 2d–f and Extended Data Fig. 7 are normalized to the effect for Chromosome 1. Precise definitions for the model, and for the 14 chromosome effect predictors tested, are given in Supplementary Information.

**Analysis code availability and source data.** Analysis code used for analysis in this study is available at <https://github.com/anjali-hinch/hybrid-rescue>. The source data generated in this publication has been deposited in NCBI's Gene Expression Omnibus (accession number GSE73833).

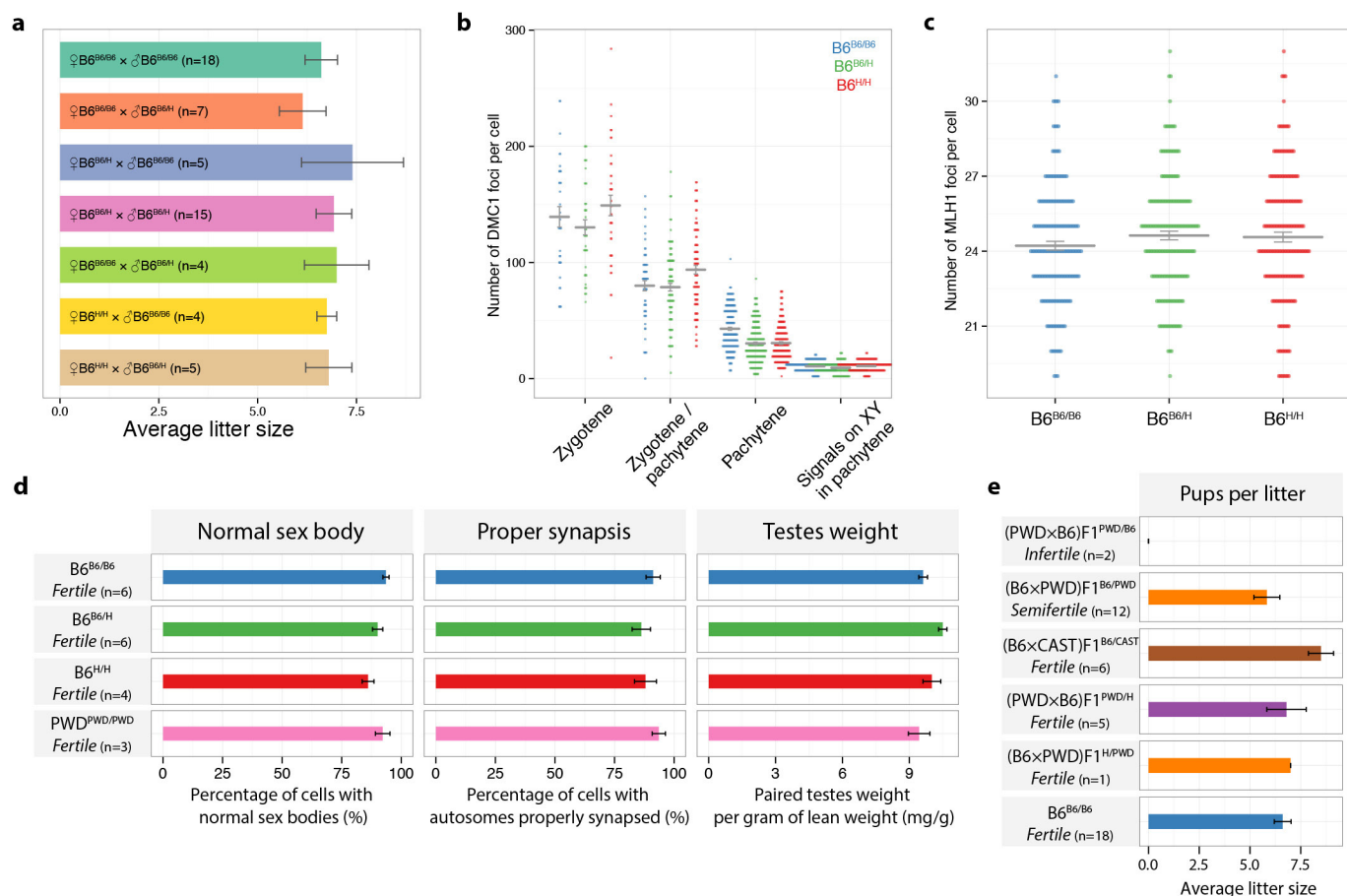
- Chen, C. M., Krohn, J., Bhattacharya, S. & Davies, B. A comparison of exogenous promoter activity at the ROSA26 locus using a  $\Phi$ C31 integrase mediated cassette exchange approach in mouse ES cells. *PLoS ONE* **6**, e23376 (2011).
- Daniel, K. *et al.* Meiotic homologue alignment and its quality surveillance are controlled by mouse HORMAD1. *Nature Cell Biol.* **13**, 599–610 (2011).
- Barchi, M. *et al.* ATM promotes the obligate XY crossover and both crossover control and chromosome axis integrity on autosomes. *PLoS Genet.* **4**, e1000076 (2008).
- Peters, A. H., Plug, A. W., van Vugt, M. J. & de Boer, P. A drying-down technique for the spreading of mammalian meiocytes from the male and female germline. *Chromosome Res.* **5**, 66–68 (1997).
- Anderson, L. K., Reeves, A., Webb, L. M. & Ashley, T. Distribution of crossing over on mouse synaptonemal complexes using immunofluorescent localization of MLH1 protein. *Genetics* **151**, 1569–1579 (1999).
- Hinch, A. G., Altemose, N., Noor, N., Donnelly, P. & Myers, S. R. Recombination in the human pseudoautosomal region PAR1. *PLoS Genet.* **10**, e1004503 (2014).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- Lunter, G. & Goodson, M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* **21**, 936–939 (2011).
- Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
- Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).





**Extended Data Figure 1 | Humanization of the zinc-finger domain of *Prdm9*.** **a**, Top, the targeting vector used for the humanization of the zinc-finger array encoded by a portion of exon 10. Middle, wild-type *Prdm9* allele. Bottom, the targeted humanized allele, following the action of Flp recombinase which removes the FRT-flanked neomycin selection cassette. The positions of primers used for the exon spanning RT-PCR are shown along with the sizes of the predicted amplification products from cDNA.

**b**, RT-PCR analysis using the exon spanning primers shown in **a** from testis cDNA prepared from wild-type (B6<sup>B6/B6</sup>) and heterozygous humanized (B6<sup>B6/H</sup>) mice. For gel source data, see Supplementary Fig. 1. **c**, Relative expression of the *Prdm9* transcript from testis cDNA prepared from wild-type (B6<sup>B6/B6</sup>), heterozygous (B6<sup>B6/H</sup>) and humanized (B6<sup>H/H</sup>) testis cDNA, normalized to *Hprt* ( $n = 2$  for each genotype). Error bars,  $\pm 1$  s.e.



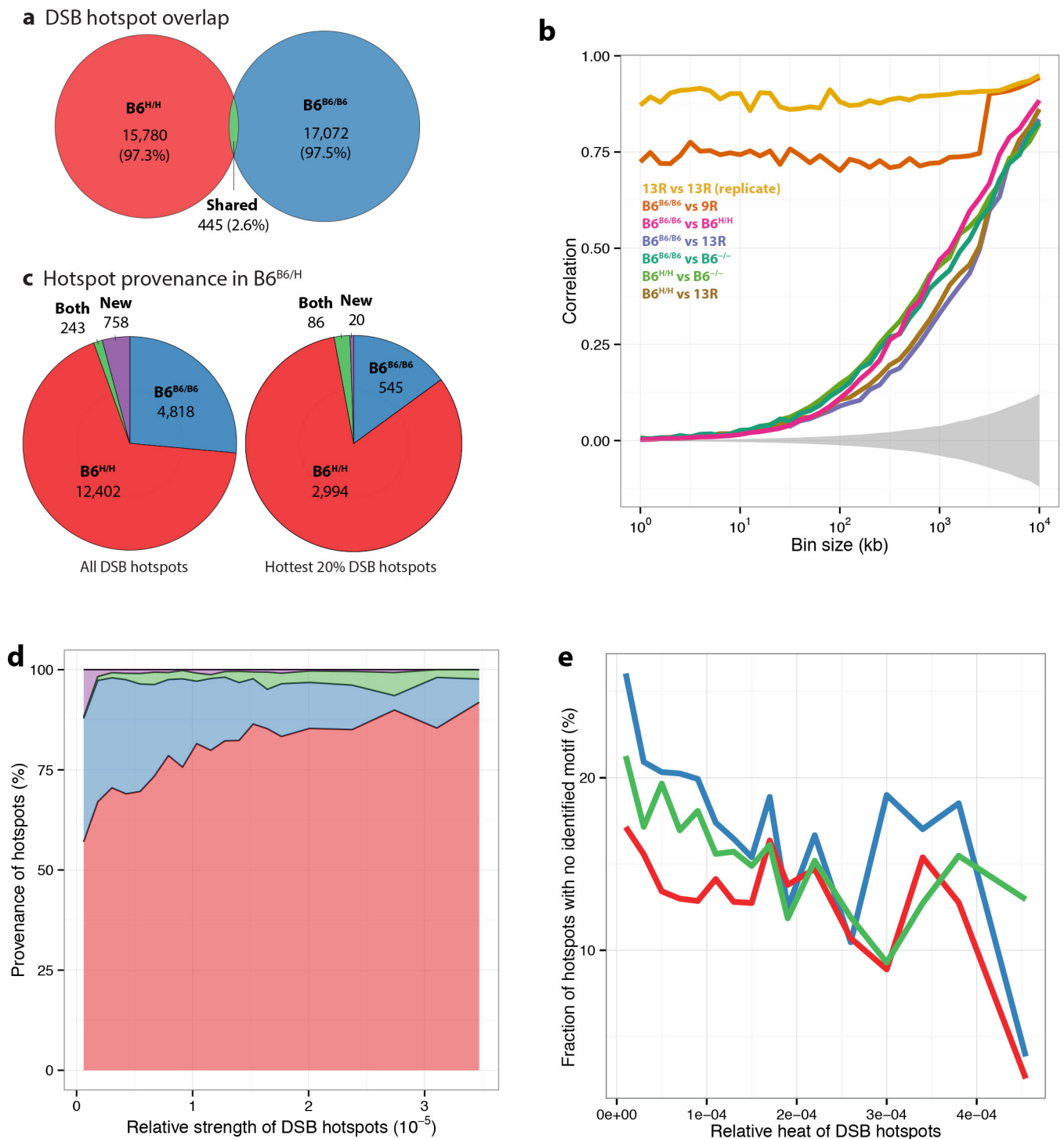
# Extended Data Figure 2 | Effects of the humanization of the *Prdm9*

zinc-finger domain on fertility parameters. **a**, The average litter size is shown for all combinations of genotype matings. Error bars,  $\pm 1$  s.e.

**b**, Numbers of DMC1 foci co-localizing with SYCP3 immunoreactivity per cell, grouped according to meiotic stage (wild-type ( $\text{B6}^{\text{B6/B6}}$ ),  $n = 5$  mice; heterozygous ( $\text{B6}^{\text{B6/H}}$ ),  $n = 7$  mice; homozygous ( $\text{B6}^{\text{H/H}}$ ),  $n = 6$  mice; cell numbers counted: zygotene: 32, 38, 37; zygotene/pachytene: 55, 96, 90; pachytene: 188, 210, 176; signals on XY in pachytene: 188, 210, 175 for

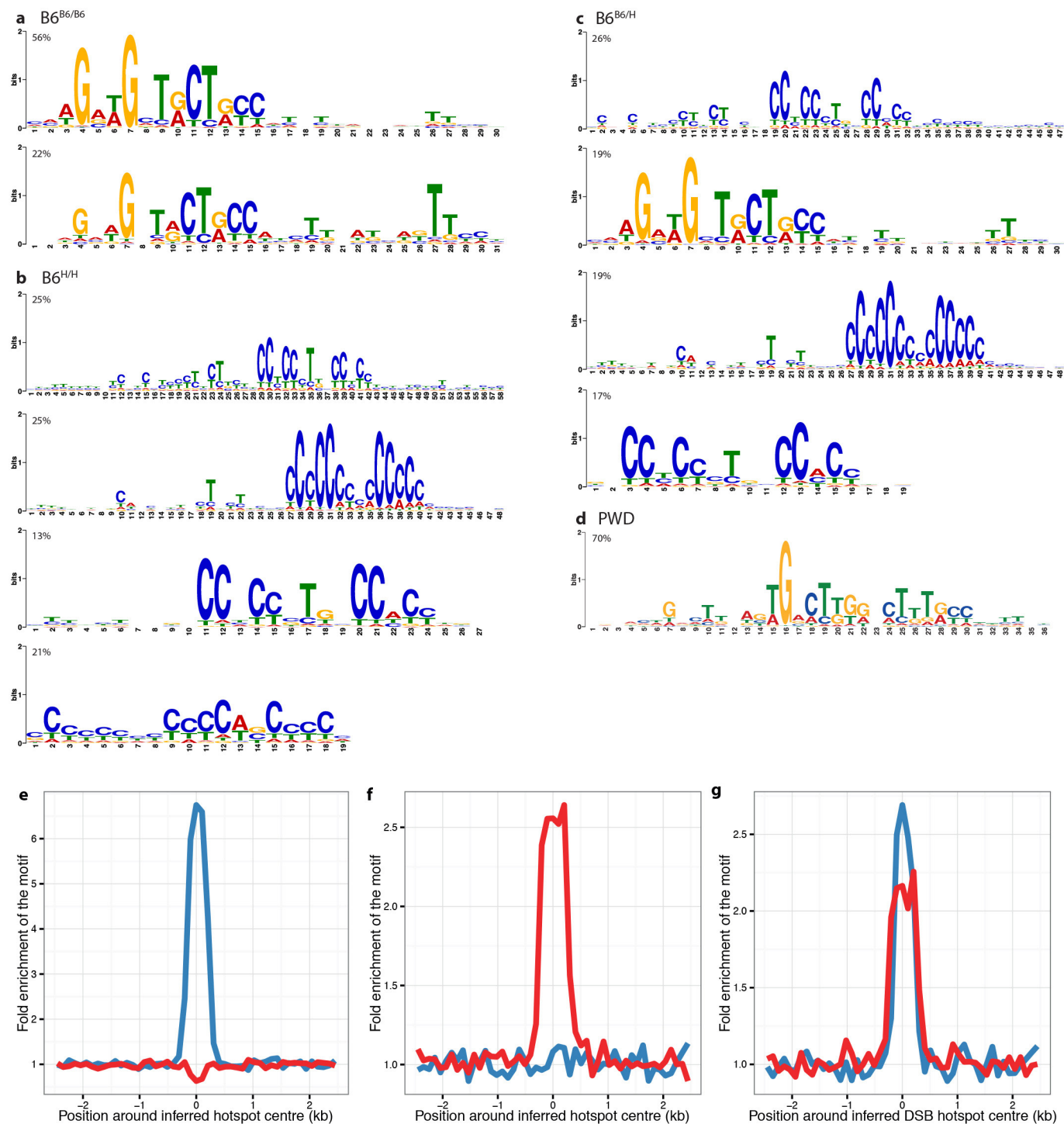
$\text{B6}^{\text{B6/B6}}$ ,  $\text{B6}^{\text{B6/H}}$  and  $\text{B6}^{\text{H/H}}$ , respectively). Mean values are shown **c**, Number of MLH1 foci per cell in pachytene stage meiotic spreads. ( $\text{B6}^{\text{B6/B6}}$ :  $n = 6$  mice, 180 cells;  $\text{B6}^{\text{B6/H}}$ :  $n = 6$  mice, 185 cells;  $\text{B6}^{\text{H/H}}$ :  $n = 6$  mice, 183 cells). Mean values are shown. **d**, Comparison of fertility metrics in four mice with homozygous genetic background (B6 or PWD). Across all four mice, there is no statistically significant evidence of differences in these fertility parameters (ANOVA, Bonferroni-corrected  $P$  values  $> 0.08$ ). Error bars,  $\pm 1$  s.e. **e**, Average litter sizes in F1 crosses. Error bars,  $\pm 1$  s.e.





**Extended Data Figure 3 | Further features revealed by DMC1 signal analysis in mice with homozygous genetic background.** **a**, Effect of humanization of the *Prdm9* zinc-finger domain on DSB hotspots. A total of 16,225 and 17,517 DSB hotspots were localized in the homozygous humanized and wild-type mice, respectively. Only 2.6% of these hotspots overlap. **b**, Correlations between DSB hotspot maps at different scales. Autosomes are divided into bins of given length, and correlations between the sums of the heats of the hotspots falling into each bin are reported, for different bin sizes. The grey region indicates the empirical 95% confidence envelope for the correlation under the null hypothesis of no association between the B6<sup>B6/B6</sup> and B6<sup>H/H</sup> DSB maps. DSB maps for B6<sup>B6/B6</sup>, 13R, 9R and *Prdm9* knockout (B6<sup>-/-</sup>) mice come from ref. 21. B6<sup>B6/B6</sup> and 9R have the same *Prdm9* allele, but different genomic backgrounds. **c**, Breakdown of

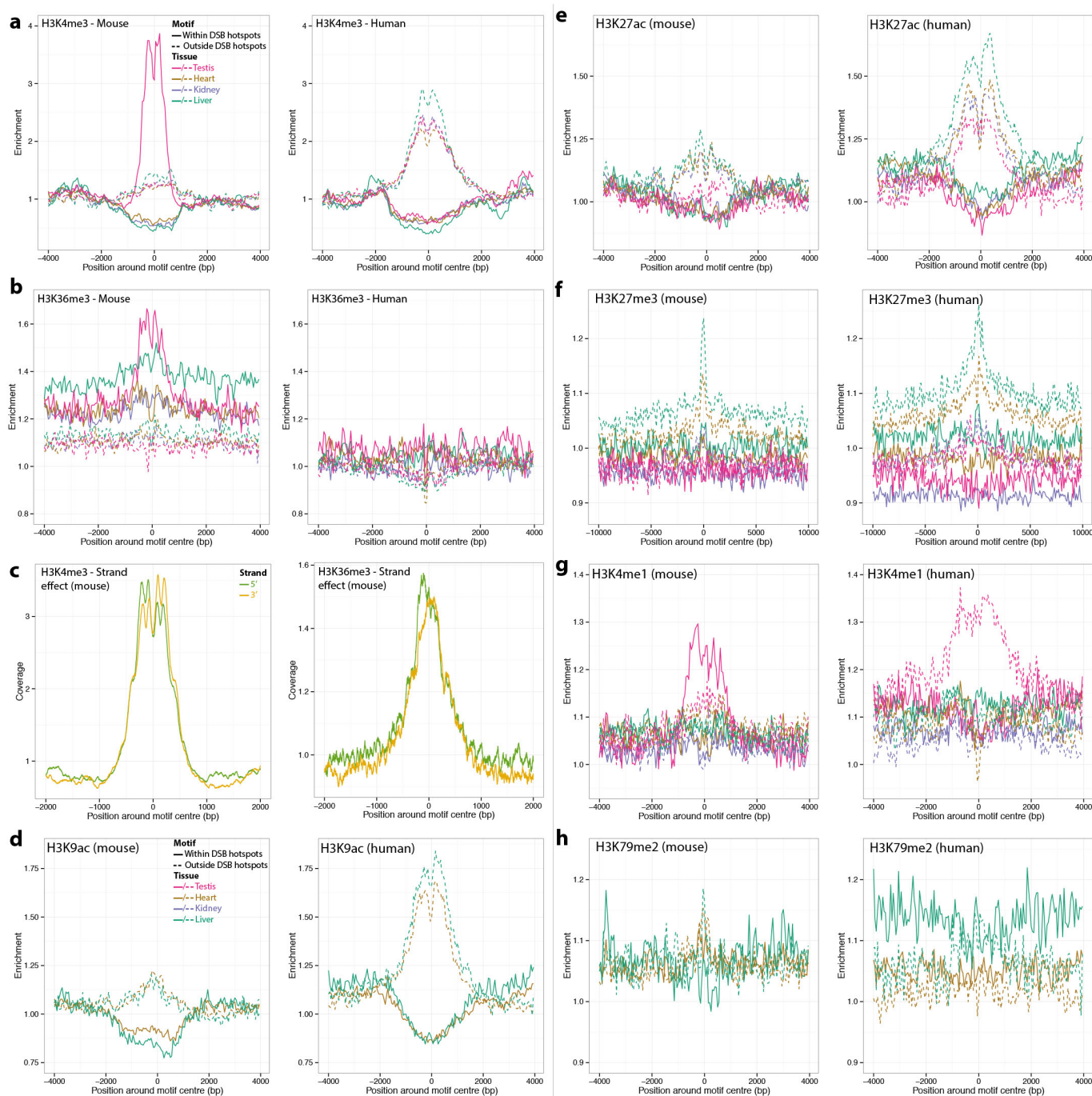
hotspot provenance (defined by overlap) in the heterozygous humanized mouse for all DSB hotspots (left panel) and for the hottest 20% of hotspots (right panel). **d**, Distributions of hotspot provenance in the heterozygous humanized mice as a function of the estimated hotspot heats (blue, wild-type B6 mouse; red, humanized homozygous mouse; green, humanized heterozygous mouse; purple, undetermined). The human allele dominates over the mouse allele in terms of heat, as the proportion of DSB hotspots found in the heterozygous mouse that are shared with the homozygous humanized mouse increases with estimated heat. The relative heat/strength of a hotspot is the ratio of this hotspot's estimated heat to the sum of all the estimated heats (on autosomes). **e**, Hotter hotspots present a PRDM9 binding motif more often than weaker hotspots in all samples (same colour legend).



**Extended Data Figure 4 | Inferred PRDM9 binding motifs are enriched at DSB hotspot centres.** a–d, Refined PRDM9 binding motifs detected in the wild-type B6 mouse (a), in the homozygous humanized mouse (b), in the heterozygous humanized mouse (c) and in wild-type PWD (d). Percentages above each motif indicate the fraction of DSB hotspots that are found to harbour this motif, with each DSB hotspot assigned at most

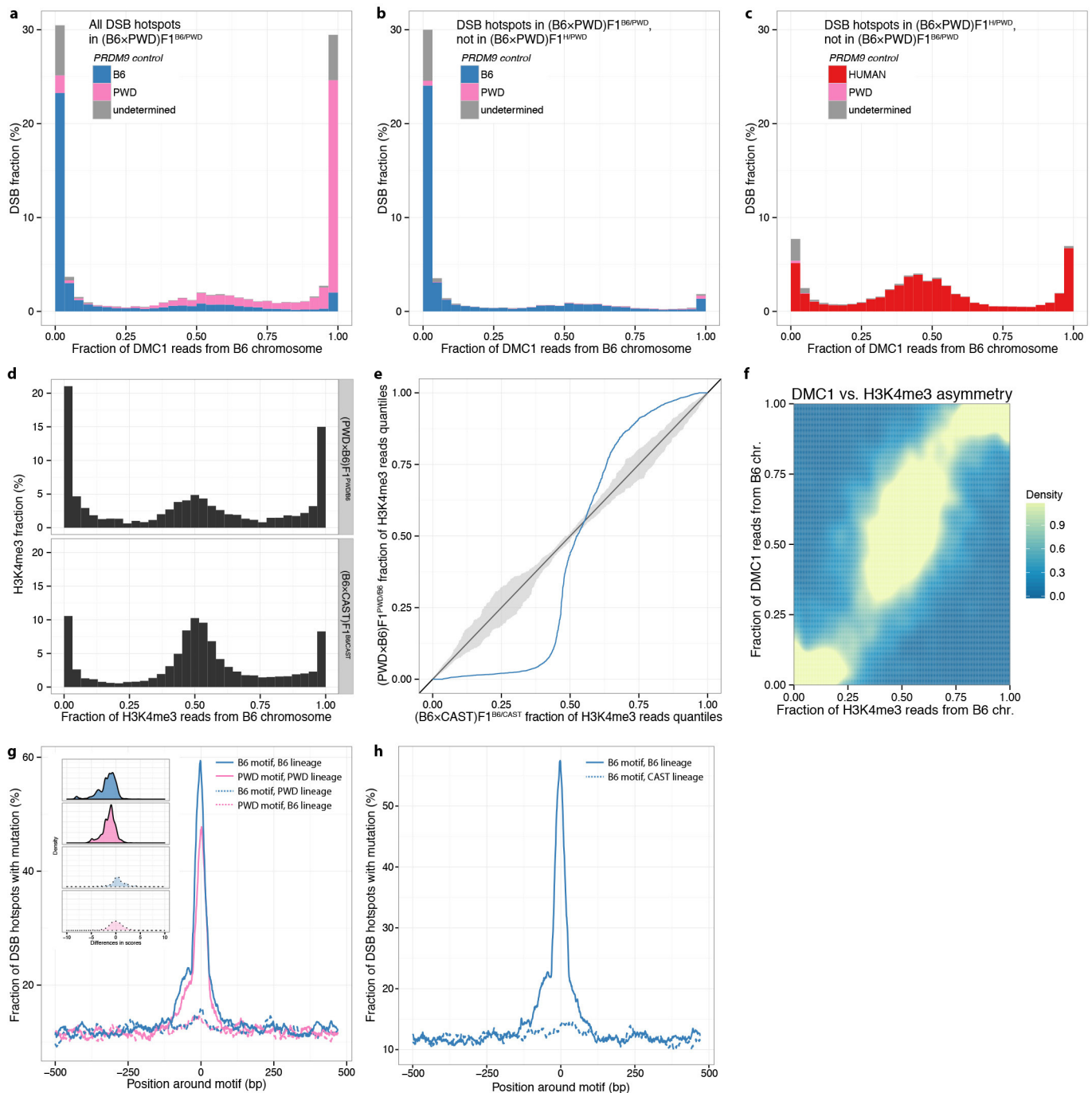
to one motif. In logo plots, letter height in bits of information determines degree of base specificity. e–g, Enrichment of the most prevalent 15-bp wild-type (blue) and humanized homozygous (red) motifs within 100-bp bins across a 5-kb window centred on the DSB hotspot centres. Enrichments were computed for the wild-type (e), homozygous humanized (f) and heterozygous humanized (g) mice DSB hotspots.





**Extended Data Figure 5 | Differential epigenetic mark distributions at PRDM9 binding motifs.** **a**, Enrichment of H3K4me3 marks at mouse motifs that are either within a B6 (left) or human (right) *PRDM9* allele controlled DSB hotspot, or outside such a hotspot. The enrichment is relative to a control genomic track. Given the spread of the distributions, the interaction range between the histones and the DSB hotspot seems to be ~1.5 kb on each side of the motif. **b**, As **a**, but for H3K36me3 marks.

**c**, Mean coverage of H3K4me3 (left) or H3K36me3 (right) signal around the mouse motif nearest to each B6 DSB hotspot, split according to the strand on which the motif lies. **d–h**, As **a**, but for H3K9ac (**d**), H3K27ac (**e**), H3K27me3 (**f**), H3K4me1 (**g**) or H3K79me2 (**h**) marks. All ChIP-seq data for histone modifications used in this analysis were obtained from the Mouse ENCODE Project.

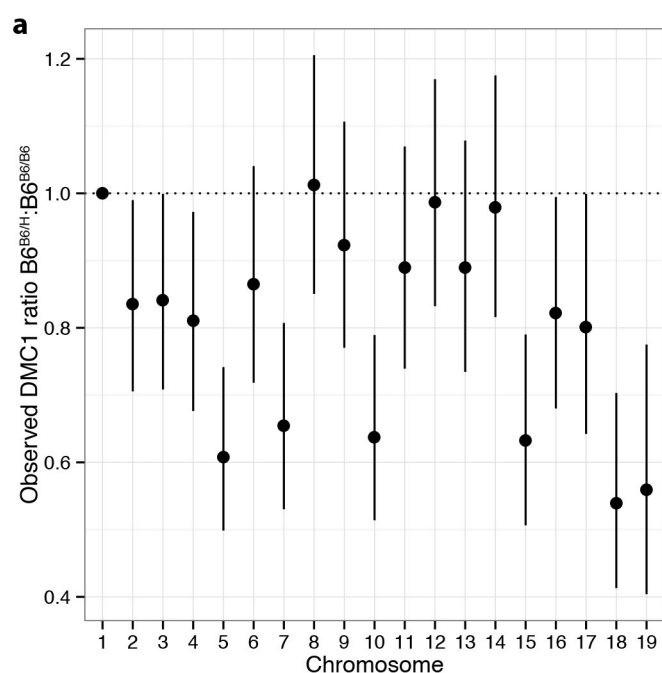


### Extended Data Figure 6 | Further features of DSB hotspot asymmetry.

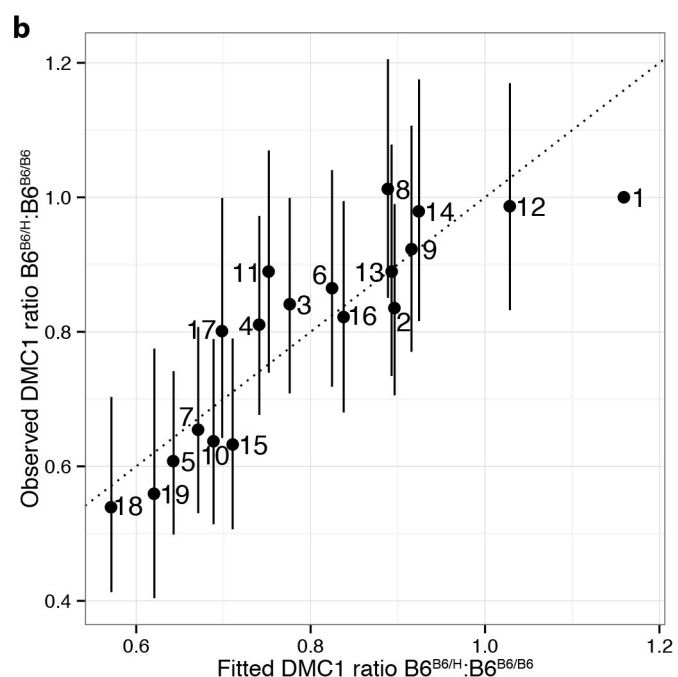
**a–c**, DSB hotspot asymmetry in  $(B6 \times PWD)F1^{B6/PWD}$  and in  $(B6 \times PWD)F1^{H/PWD}$ . **a**, Distribution of the fraction of (DMC1) informative reads originating from the B6 chromosome in the reciprocal  $(B6 \times PWD)F1^{B6/PWD}$  mouse. PRDM9 control at each DSB is attributed either to the B6 allele (blue) or the PWD allele (pink) or is undeterminable (grey). **b**, **c**, As **a**, but showing fractions only for non-shared hotspots, unique to either the reciprocal  $(B6 \times PWD)F1^{B6/PWD}$  (**b**) or the reciprocal rescue  $(B6 \times PWD)F1^{H/PWD}$  (**c**) mice. **d**, **e**, Comparison of the levels of asymmetric binding in the  $(PWD \times B6)F1^{PWD/B6}$  and  $(B6 \times CAST)F1^{B6/CAST}$  mice, using H3K4me3 signal. **d**, Distributions of the fraction of H3K4me3 reads from the B6 chromosome in the two mice. We used raw data from ref. 17 for the  $(B6 \times CAST)F1^{B6/CAST}$  mouse, and processed both data sets in the same way. H3K4me3 heats were capped at the 95th percentile in each case, and only H3K4me3 binding peaks not inferred to be independent of PRDM9 binding (Supplementary Information Section 7), and overlapping with a DMC1 hotspot in the same mouse, were considered. **e**, Quantile–quantile plot for the distributions shown in **d** (blue). Dark grey:  $y = x$  line; light

grey: 95% confidence band. **f**, Density plot comparing, for each hotspot in the  $(PWD \times B6)F1^{PWD/B6}$  mouse, its DMC1 and H3K4me3 asymmetries. The correlation between the two measures is 0.93. **g**, Mutations within 1-kb regions around B6 and PWD PRDM9 motifs, on the B6 and PWD genomes. Main plot, for each combination of motif and lineage (PWD or B6), we plot the fraction of 30-bp windows, along the 1-kb regions surrounding motif occurrences within DSB hotspots, where at least one SNP or indel mutation occurred along the respective lineage. Inset plot, distribution of motif score differences (derived-ancestral) for motif changes shown in the main plot. Motif score was defined as the logarithm of the probability that a motif was drawn from the motif's position weight matrix, in the ancestral sequence and in the current-day mouse. A negative difference indicates the motif match worsened along the corresponding lineage. This panel is based on the  $(PWD \times B6)F1^{PWD/B6}$  DMC1 map. **h**, Mutations within 1-kb regions around B6 PRDM9 motifs, on the B6 and CAST genomes, as in **g**, using the  $(B6 \times CAST)F1^{B6/CAST}$  DMC1 map. We see no evidence of erosion of B6 PRDM9 motifs on the CAST genome.

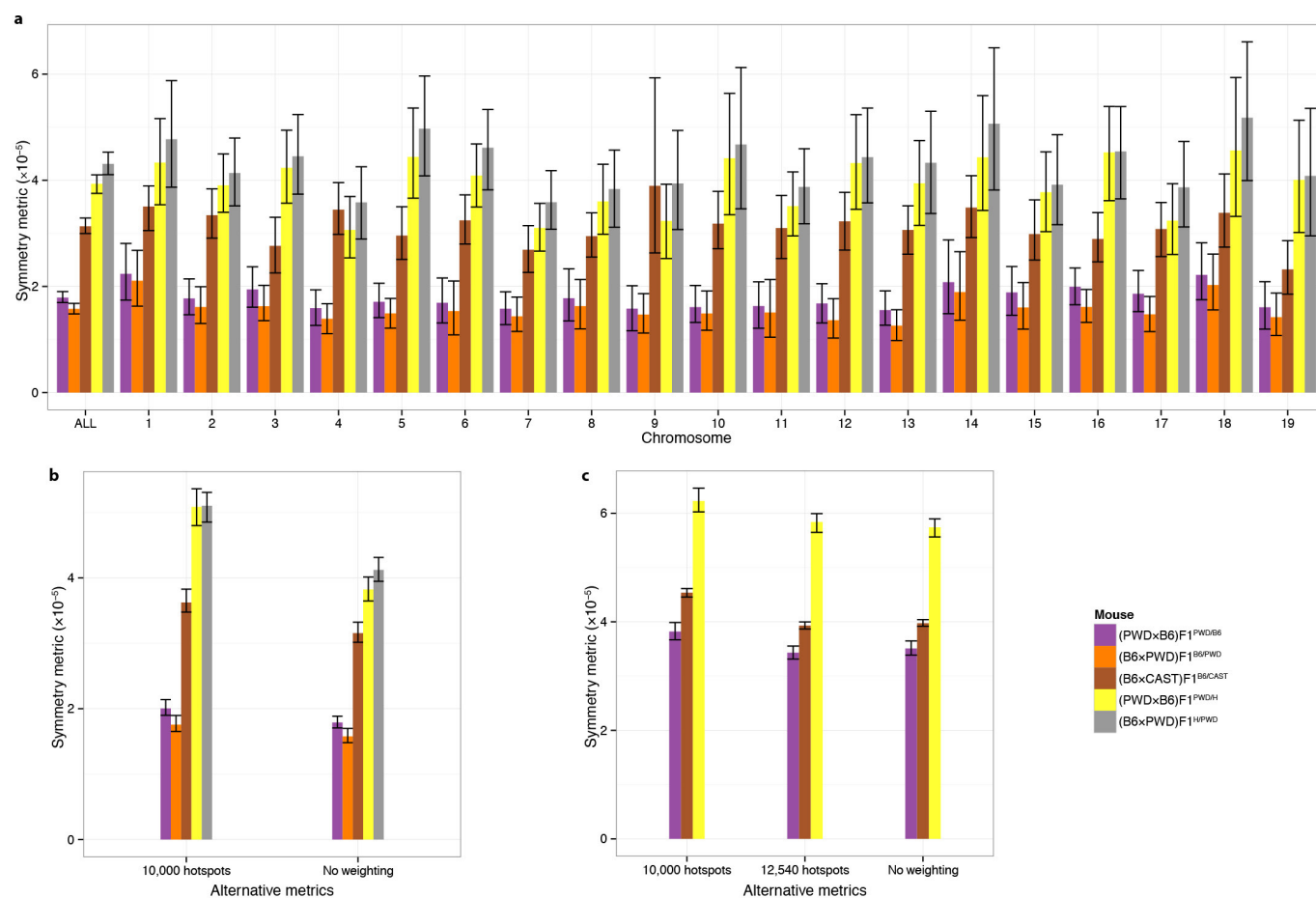




**Extended Data Figure 7 | Chromosome effects following *Prdm9* humanization in  $B6^{B6/B6}$ .** **a**, Individual chromosome effects (relative to chromosome 1) when comparing DMC1 signals in the  $B6^{B6/H^H}$  mouse relative to the  $B6^{B6/B6}$  mouse, for the DSB hotspots that are shared between



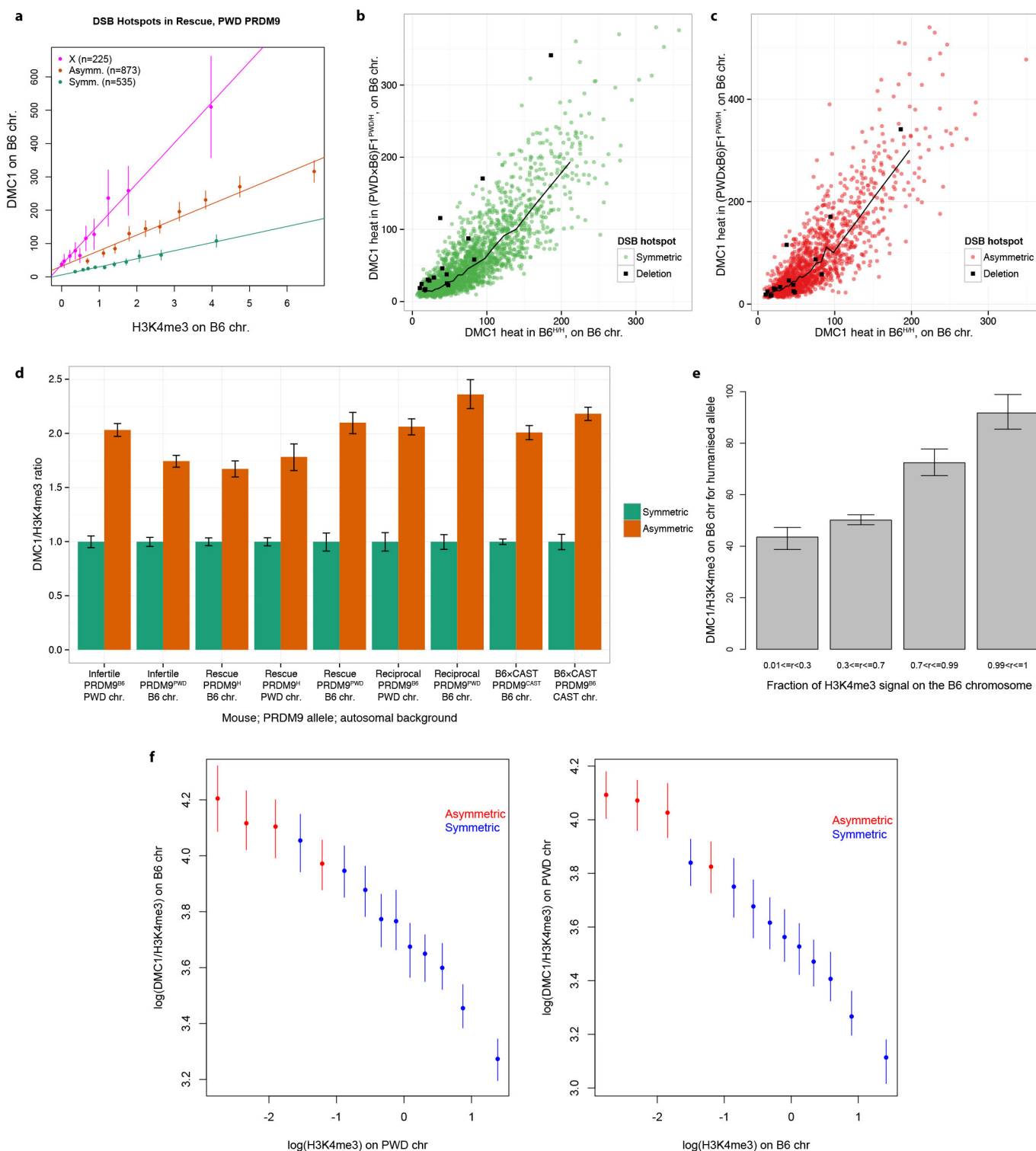
these two mice. **b**, Comparison of the observed chromosome effects for DMC1 signals with the fitted chromosome effects, using the two-predictor model including the sum of symmetric H3K4me3 heats in  $B6^{B6/B6}$  and in  $B6^{H/H}$ . Bars conservatively show 3 standard errors in both plots.



**Extended Data Figure 8 | Value by chromosome and sensitivity analysis for the symmetry metric.** **a**, Symmetry metric, as defined in the main text, for each sample (ALL), and for each autosome amongst those samples. Error bars represent bootstrap 95% confidence intervals in all panels. **b**, Alternative symmetric metrics (to the ones reported in the main text), using only 10,000 hotspots per sample, or without weighting each

chromosome specific metric, to compute the average metric genome-wide. Both metrics are computed using the DMC1 maps. **c**, Alternative symmetric metrics using H3K4me3 maps, similarly to **b**. The threshold of 12,540 hotspots per sample corresponds to the number of hotspots with ratio estimates in the (PWD  $\times$  B6)F1<sup>PWD/H</sup> mouse, which was the lowest amongst the three samples shown here.

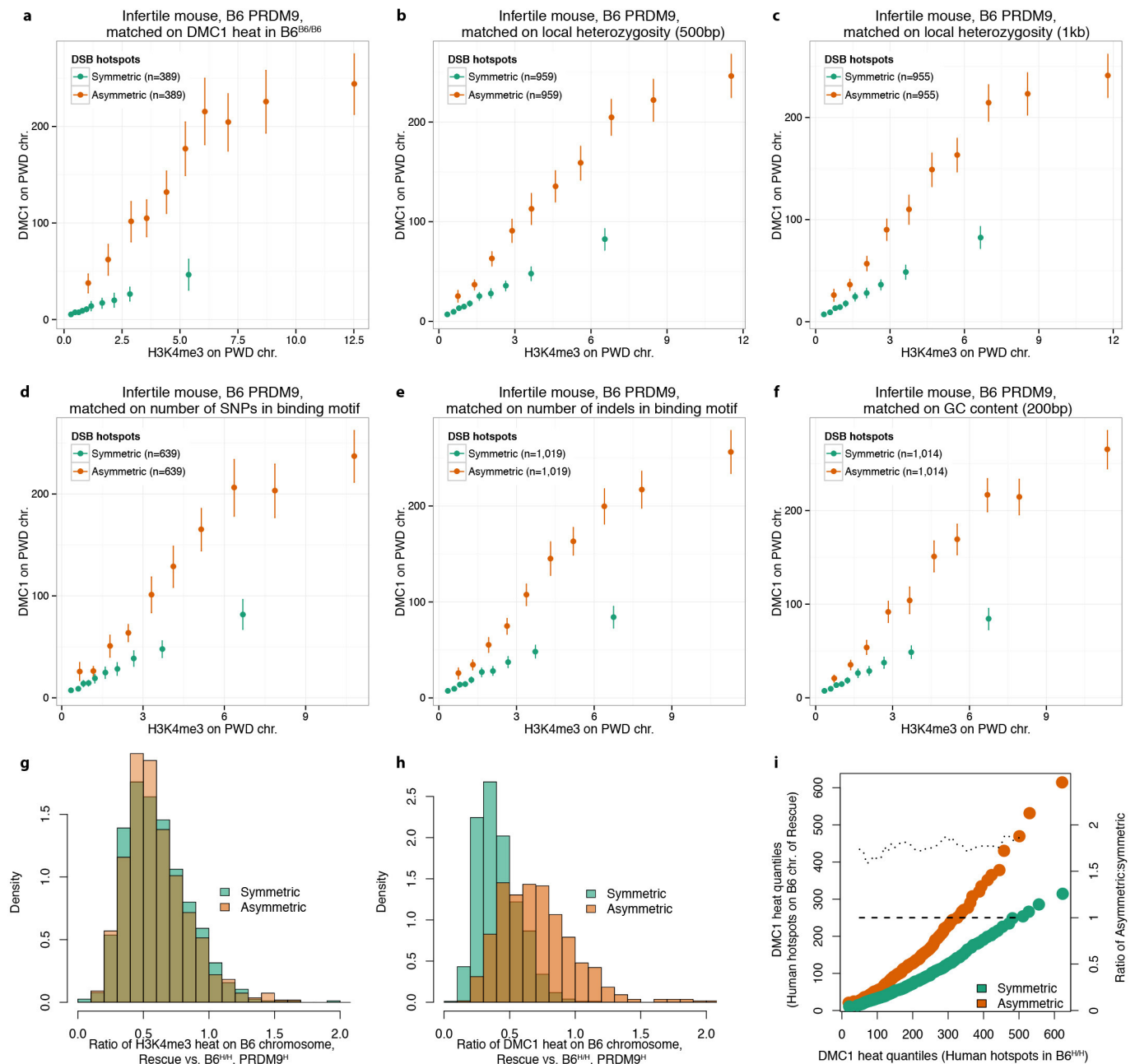




Extended Data Figure 9 | See next page for figure caption.

**Extended Data Figure 9 | Asymmetric hotspots, hotspots on the X chromosome and hotspots opposite deletions show systematic increase of DMC1 heat, relative to symmetric hotspots.** **a**, For the PWD allele in the humanized rescue (PWD  $\times$  B6)F1<sup>PWD/H</sup> mouse, mean DMC1 signal is plotted in decile bins of H3K4me3 enrichment on the B6 chromosome (or the PWD X chromosome), with error bars showing 95% confidence intervals and lines of best fit (as in Fig. 4c). The slope of the line for asymmetric hotspots is 2.5-fold greater than that of the symmetric hotspots, and the slope for hotspots on the X chromosome is 5.2-fold greater, illustrating that the DMC1 signal at asymmetric sites is elevated in a similar fashion to hotspots on the X chromosome, which do not repair until late in meiosis. We found similar results in all cases tested. **b**, Comparison of DMC1 heats on B6 chromosome for hotspots shared between the humanized B6<sup>H/H</sup> and the humanized rescue (PWD  $\times$  B6)F1<sup>PWD/H</sup> mice, under humanized PRDM9 control. We show symmetric hotspots (fraction of DMC1 informative reads between 0.4 and 0.6, green), and hotspots opposite a deletion on the PWD chromosome (deletion of at least 200 bp, encompassing a human PRDM9 binding motif, black). The black line represents the median DMC1 heat for symmetric hotspots. **c**, As **b**, but showing the asymmetric hotspots (fraction of DMC1 informative reads above 0.9, red), with the corresponding median line. Hotspots opposite PWD deletion show a significant elevation in DMC1 heat relative to symmetric hotspots (14 of 16 hotspots above the symmetric median line,  $P = 0.004$ ). This elevation is similar to the one showed by asymmetric hotspots (9 of 16 hotspots above the asymmetric median line,  $P = 0.80$ ). **d**, Bar plot showing the genome-wide autosomal ratio of mean DMC1 heat to mean H3K4me3 enrichment for asymmetric hotspots relative to symmetric hotspots in 9 scenarios studied, each for a different combination of mouse, *Prdm9* allele, and haplotype, with error bars representing 95% bootstrap confidence intervals for the ratio

of means. In all cases, asymmetric hotspots show an elevation in DMC1 signal for a given H3K4me3 signal. **e**, Ratio of mean DMC1 and H3K4me3 signals on the B6 chromosome for the humanized allele in the humanized rescue mouse. Hotspots are clustered according to the fractions of their H3K4me3 signal that is on the B6 chromosome ( $r$ ), and the ratio of the mean DMC1 and H3K4me3 signals in each class is shown here. The whiskers show 95% confidence intervals for the mean, estimated using bootstrapping. When  $r > 0.5$ , the B6 chromosome has greater H3K4me3 than the PWD chromosome, and vice versa. The ratio could not be estimated for  $r \leq 0.01$  owing to H3K4me3 levels being zero or nearly zero in those cases. **f**, Left, ratio of mean DMC1 and H3K4me3 signals on the B6 chromosome compared with the H3K4me3 signal on the PWD chromosome (log scale) in the infertile mouse. Asymmetric hotspots were defined as those with H3K4me3 fraction on the B6 chromosome  $> 0.9$ , and symmetric hotspots were those with the fraction between 0.1 and 0.9. Hotspots that we estimated to be completely asymmetric (H3K4me3 fraction = 0 on either chromosome) or those with H3K4me3 enrichment on either chromosome close to zero (enrichment  $< 0.05$ ) were excluded to avoid singularities on either axis. Asymmetric hotspots were binned into four bins of equal size and symmetric hotspots were binned into ten bins of equal size. Different numbers of bins were used for asymmetric and symmetric hotspots to get approximately similar confidence intervals (error bars represent 95% confidence intervals) to enable comparison. We did not observe many weak symmetric hotspots as we have limited power to detect such hotspots, which is why there are no symmetric bins with very low H3K4me3 levels on the homologue (right). Right, as left, but with the ratio determined for the PWD chromosome relative to H3K4me3 on the B6 chromosome. Accordingly, asymmetric hotspots are defined as those with H3K4me3 fraction on the PWD chromosome  $> 0.9$ .



**Extended Data Figure 10 | Elevation of DMC1 asymmetric heat is not explained by GC content, local heterozygosity, differences in binding-motif-disrupting mutations or by outliers. a–f,** Comparison of DMC1 signals in the infertile (PWD  $\times$  B6)F1<sup>PRDM9/B6</sup> mouse, at symmetric and asymmetric hotspots respectively, binned by H3K4me3 enrichment, after matching symmetric and asymmetric hotspots on various features: **a**, DMC1 heat in B6<sup>B6/B6</sup>; **b**, local heterozygosity outside the PRDM9 binding motif, in a 500-bp window; **c**, as **b**, but for a 1-kb window; **d**, number of SNPs in binding motif; **e**, number of indels in binding motif; **f**, local GC content, computed in a 200-bp window around hotspot centre. **g**, Distributions of the ratios of H3K4me3 heats on the B6 chromosome, in the rescue (PWD  $\times$  B6)F1<sup>PRDM9/H</sup> versus humanized B6<sup>H/H</sup> mice, for the symmetric (fraction of informative DMC1 reads in the range 0.4–0.6, green) and asymmetric (fraction 0.9–1, orange) hotspots under humanized PRDM9 control shared between the two mice. The distributions are very close, suggesting similar trimethylation by PRDM9 on the B6 chromosome

in both mice. **h**, As **g**, but for the DMC1 heats. Despite similar trimethylation marking by PRDM9 in both mice, we observed notable changes in the distribution of DMC1 ratios. This could be due to either more breaks occurring at the asymmetric sites, or a longer time taken to repair them. **i**, Quantile–quantile plots of DMC1 heats for hotspots under the control of the humanized allele on the B6 chromosome in the rescue (PWD  $\times$  B6)F1<sup>PRDM9/H</sup> (*y* axis, left) versus the humanized B6<sup>H/H</sup> (*x* axis), for symmetric (green) and asymmetric (orange) hotspots. Dotted line represents the ratios of asymmetric to symmetric quantiles (excluding distribution tails; *y* axis, right). Dashed line represents expected ratio if there were no differences between symmetric and asymmetric hotspots. The observed ratio of DMC1 quantiles is constant across DMC1 heats, emphasizing that the increase in DMC1 heat at asymmetric sites is very similar across the whole range of DMC1 heats, and does not simply result from a few outlying hotspots.



# Schizophrenia risk from complex variation of complement component 4

Aswin Sekar<sup>1,2,3</sup>, Allison R. Bialas<sup>4,5</sup>, Heather de Rivera<sup>1,2</sup>, Avery Davis<sup>1,2</sup>, Timothy R. Hammond<sup>4</sup>, Nolan Kamitaki<sup>1,2</sup>, Katherine Tooley<sup>1,2</sup>, Jessy Presumey<sup>5</sup>, Matthew Baum<sup>1,2,3,4</sup>, Vanessa Van Doren<sup>1</sup>, Giulio Genovese<sup>1,2</sup>, Samuel A. Rose<sup>2</sup>, Robert E. Handsaker<sup>1,2</sup>, Schizophrenia Working Group of the Psychiatric Genomics Consortium\*, Mark J. Daly<sup>2,6</sup>, Michael C. Carroll<sup>5</sup>, Beth Stevens<sup>2,4</sup> & Steven A. McCarroll<sup>1,2</sup>

**Schizophrenia is a heritable brain illness with unknown pathogenic mechanisms. Schizophrenia's strongest genetic association at a population level involves variation in the major histocompatibility complex (MHC) locus, but the genes and molecular mechanisms accounting for this have been challenging to identify. Here we show that this association arises in part from many structurally diverse alleles of the complement component 4 (C4) genes. We found that these alleles generated widely varying levels of C4A and C4B expression in the brain, with each common C4 allele associating with schizophrenia in proportion to its tendency to generate greater expression of C4A. Human C4 protein localized to neuronal synapses, dendrites, axons, and cell bodies. In mice, C4 mediated synapse elimination during postnatal development. These results implicate excessive complement activity in the development of schizophrenia and may help explain the reduced numbers of synapses in the brains of individuals with schizophrenia.**

Schizophrenia is a heritable psychiatric disorder involving impairments in cognition, perception, and motivation that usually manifest late in adolescence or early in adulthood. The pathogenic mechanisms underlying schizophrenia are unknown, but observers have repeatedly noted pathological features involving excessive loss of grey matter<sup>1,2</sup>, and reduced numbers of synaptic structures on neurons<sup>3–5</sup>. Although treatments exist for the psychotic symptoms of schizophrenia, there is no mechanistic understanding of, nor effective therapies to prevent or treat, the cognitive impairments and deficit symptoms of schizophrenia, which are the earliest and most constant features of the disorder. An important goal in human genetics is to find the biological processes that underlie such disorders.

More than 100 loci in the human genome contain single nucleotide polymorphism (SNP) haplotypes that associate with risk of schizophrenia<sup>6</sup>; however, the functional alleles and mechanisms at these loci remain to be discovered. By far the strongest such genetic relationship is schizophrenia's association with genetic markers across the major histocompatibility complex (MHC) locus, which spans several megabases (Mb) of chromosome 6 (refs 6–10). The MHC locus is best known for its role in immunity, containing 18 highly polymorphic human leukocyte antigen (*HLA*) genes that encode a vast suite of antigen-presenting molecules. In some autoimmune diseases, genetic associations at the MHC locus arise from alleles of *HLA* genes<sup>11,12</sup>; however, schizophrenia's association to the MHC has not yet been explained.

Though the functional alleles that give rise to genetic associations have in general been challenging to find, the schizophrenia–MHC association has been particularly challenging because schizophrenia's complex pattern of association to markers in the MHC locus spans hundreds of genes and does not correspond to the linkage disequilibrium (LD) around any known variant<sup>6,10</sup>. This prompted us to consider cryptic genetic influences that might generate unconventional genetic signals. The most strongly associated markers in

several large case/control cohorts were near a complex, multi-allelic, and only partially characterized form of genome variation that affects the *C4* gene encoding complement component 4 (Extended Data Fig. 1). The association of schizophrenia to *CSMD1* (refs 6, 10), which encodes a regulator of C4 (ref. 13), further motivated us to consider *C4*.

## C4 structures and MHC SNP haplotypes

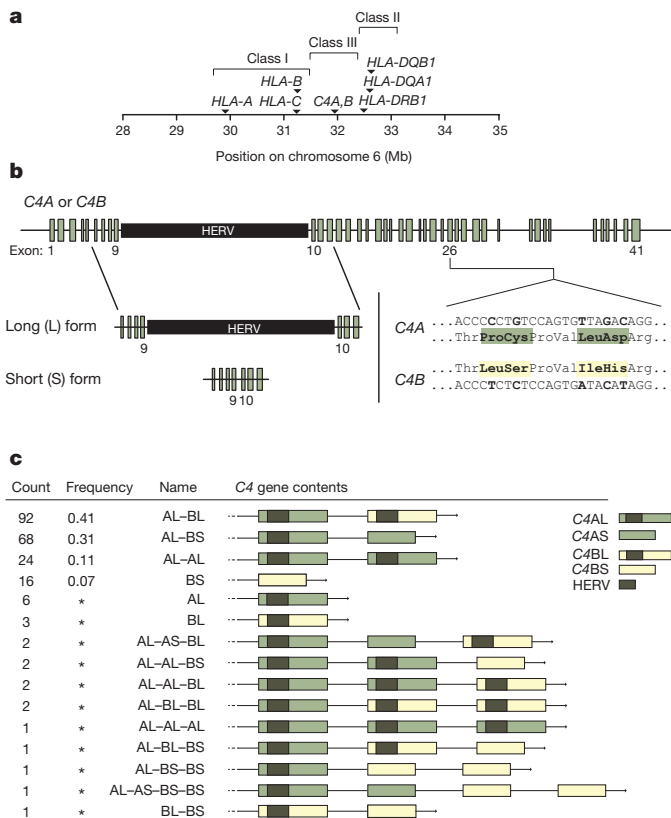
Human *C4* exists as two functionally distinct genes (isotypes), *C4A* and *C4B*; both vary in structure and copy number. One to three *C4* genes (*C4A* and/or *C4B*) are commonly present as a tandem array within the MHC class III region (Fig. 1a and Extended Data Fig. 1g)<sup>14–18</sup>. The protein products of *C4A* and *C4B* bind different molecular targets<sup>19,20</sup>. *C4A* and *C4B* segregate in both long (L) and short (S) genomic forms (*C4AL*, *C4AS*, *C4BL* and *C4BS*), distinguished by the presence or absence (in intron 9) of a human endogenous retroviral (HERV) insertion that lengthens *C4* from 14 to 21 kb without changing the *C4* protein sequence<sup>16</sup> (Fig. 1b).

We developed a way (Extended Data Fig. 2) to identify the 'structural haplotypes' of *C4*—the copy number of *C4A* and *C4B* and the long/short (HERV) status of each *C4A* and *C4B* copy—present on 222 copies of human chromosome 6. Using droplet digital PCR (ddPCR), we found that genomes contained 0–5 *C4A* genes, 0–3 *C4B* genes, 1–5 long (L) *C4* genes, and 0–3 short (S) *C4* genes (Extended Data Fig. 2a, b). We also developed assays to determine the long/short (HERV) status of each *C4A* and *C4B* gene copy (Extended Data Fig. 2c), thus revealing copy number of *C4AL*, *C4BL*, *C4AS*, and *C4BS* in each genome (Supplementary Methods).

We analysed inheritance in father–mother–offspring trios (Extended Data Fig. 2d) to identify the *C4A* and *C4B* contents of individual alleles (Extended Data Fig. 2e). We found that four common *C4* structural haplotypes (AL–BL, AL–BS, AL–AL, and BS) were collectively present

<sup>1</sup>Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>2</sup>Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. <sup>3</sup>MD-PhD Program, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>4</sup>Department of Neurology, F.M. Kirby Neurobiology Center, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>5</sup>Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, Massachusetts 02115, USA. <sup>6</sup>Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.

\*Lists of participants and their affiliations appear in the Supplementary Information.



**Figure 1 | Structural variation of the complement component 4 (C4) gene.** **a**, Location of the C4 genes within the major histocompatibility complex (MHC) locus on human chromosome 6. **b**, Human C4 exists as two paralogous genes (isotypes), C4A and C4B; the encoded proteins are distinguished at a key site that determines which molecular targets they bind<sup>19,20</sup>. Both C4A and C4B also exist in both long (L) and short (S) forms distinguished by an endogenous retroviral (C4-HERV) sequence in intron 9. **c**, Structural forms of the C4 locus and their frequencies among a European-ancestry population sample (222 chromosomes from 111 genetically unrelated individuals, HapMap CEU), inferred as described in Extended Data Fig. 2. Asterisks indicate allele frequencies too low to be estimated accurately.

on 90% of the 222 independent chromosomes sampled; 11 uncommon C4 haplotypes comprised the other 10% (Fig. 1c).

The series of many SNP alleles along a genomic segment (the SNP haplotype) can be used to identify chromosomal segments that come from shared common ancestors. We identified the SNP haplotype(s) on which each C4 locus structure was present (Fig. 2). The three most common C4 locus structures were each present on multiple MHC SNP haplotypes (Fig. 2). For example, the C4 AL-BS structure (frequency 31%) was present on five common haplotypes (frequencies 4%, 4%, 4%, 8%, and 6%) and many rare haplotypes (collective frequency 5%, Fig. 2). Reflecting this haplotype diversity, each of these C4 structures exhibited real but only partial correlation to individual SNPs (Extended Data Fig. 3). The relationship between C4 structures and SNP haplotypes was generally one-to-many: a C4 structure might be present on many haplotypes, but a given SNP haplotype tended to have one characteristic C4 structure (Fig. 2).

### C4 expression variation in the brain

As C4A and C4B vary in both copy number and C4-HERV status (Fig. 1), and because other HERVs can function as enhancers<sup>21–23</sup>, C4 variation might affect expression of C4 genes. We assessed how C4 structural variation related to RNA expression of C4A and C4B in eight panels of post-mortem human adult brain samples (674 samples from 245 distinct donors in 3 cohorts, Supplementary Methods).

The results of this expression analysis were consistent across all five brain regions analysed. First, RNA expression of C4A and C4B increased proportionally with copy number of C4A and C4B respectively (Fig. 3a, b and Extended Data Fig. 4). These observations mirror earlier observations in human serum<sup>24</sup>. Second, expression levels of C4A were two to three times greater than expression levels of C4B, even after controlling for relative copy number in each genome (Fig. 3c). Third, copy number of the C4-HERV sequence increased the ratio of C4A to C4B expression ( $P < 10^{-7}$ ,  $P < 10^{-2}$ ,  $P < 10^{-3}$ , respectively, in the three cohorts examined, by Spearman rank correlation) (Fig. 3c and Extended Data Fig. 4).

We used the above data to create genetic predictors of C4A and C4B expression levels in the brain (Supplementary Methods). If C4A or C4B expression levels influence a phenotype, then the aggregate genetic predictor might associate to schizophrenia more strongly than individual variants do.

### C4 structural variation in schizophrenia

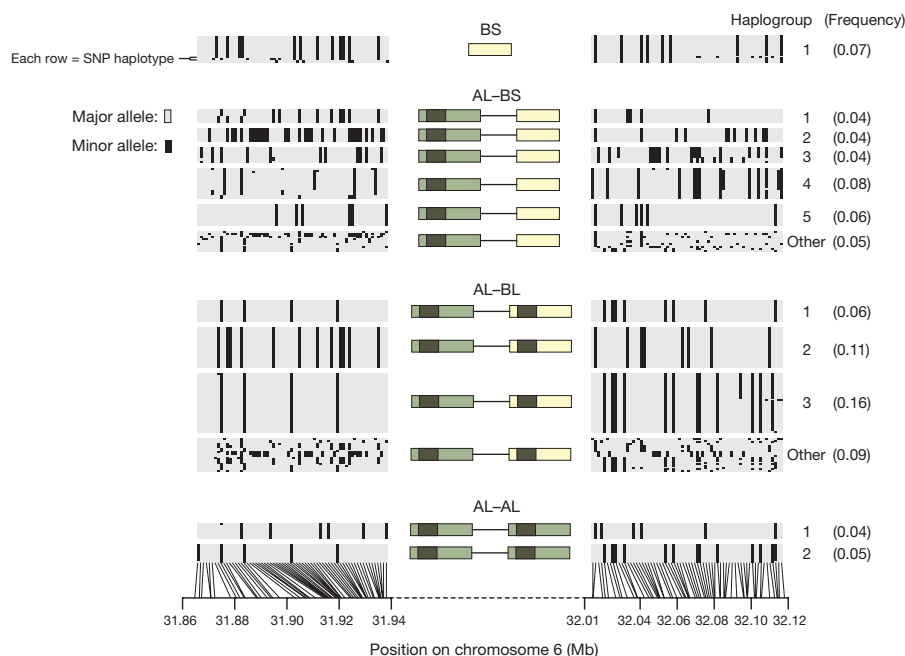
Schizophrenia cases and controls from 22 countries have been analysed genome-wide for SNPs, implicating the MHC locus as the strongest of more than 100 genome-wide-significant associations<sup>6</sup>. Our analysis indicated that long haplotypes defined by many SNPs carry characteristic C4 alleles (Fig. 2), potentially making it possible to infer C4 alleles by statistical imputation<sup>25</sup> from combinations of many SNPs. We used our 222 integrated haplotypes of MHC SNPs and C4 alleles (Fig. 2) as reference chromosomes for imputation. We found that the four most common structural forms of the C4A/C4B locus (BS, AL-BS, AL-BL, and AL-AL) could be inferred with reasonably high accuracy (generally  $0.70 < r^2 < 1.00$ , Supplementary Table 2).

We then analysed SNP data from 28,799 schizophrenia cases and 35,986 controls, from 40 cohorts in 22 countries contributing to the Psychiatric Genomics Consortium (PGC)<sup>6</sup>. We evaluated association to 7,751 SNPs across the extended MHC locus (chr6: 25–34 Mb), to C4 structural alleles (Fig. 1c), and to HLA sequence polymorphisms imputed from the SNP data. We also predicted levels of C4A and C4B expression from the imputed C4 structural alleles.

The association of schizophrenia to these genetic variants exhibited two prominent features (Fig. 4a, b). One feature involved a large set of similarly associating SNPs spanning 2 Mb across the distal end of the extended MHC region (we use this set's most strongly associating SNP, rs13194504, as its genetic proxy). The other peak of association centred at C4, where schizophrenia associated most strongly with the genetic predictor of C4A expression levels ( $P = 3.6 \times 10^{-24}$ ) (Fig. 4a and Extended Data Fig. 5). In the region near C4 (chromosome 6, 31–33 Mb), the more strongly a SNP correlated with predicted C4A expression, the more strongly it associated with schizophrenia (Fig. 4b, bottom panel).

Although the variation at C4 and in the distal extended MHC region associated with schizophrenia with similar strengths ( $P = 3.6 \times 10^{-24}$  and  $5.5 \times 10^{-28}$ , respectively), their correlation with each other was low ( $r^2 = 0.18$ , Fig. 4b), suggesting that they reflect distinct genetic influences. Conditional analysis confirmed this: in analyses controlling for either rs13194504 or genetically predicted C4A expression, the other genetic variable still defined a genome-wide-significant association peak ( $P = 7.8 \times 10^{-10}$  and  $8.0 \times 10^{-14}$ , respectively, Fig. 4c, d). Controlling for both genetic variables revealed a third association signal just proximal to the MHC locus (Fig. 4e) involving SNPs around *BAK1* and *SYNGAP1*, the latter of which encodes a major component of the postsynaptic density; *de novo* loss-of-function mutations in *SYNGAP1* associate with autism<sup>26</sup>. In joint analysis, all three genetic signals remained significant ( $P = 8.0 \times 10^{-14}$ ,  $2.8 \times 10^{-8}$ , and  $1.7 \times 10^{-8}$ , respectively) and no additional genome-wide-significant signals remained in the MHC locus (Fig. 4f).

In some autoimmune diseases with genetic associations in the MHC locus, alleles of HLA genes associate more strongly than do other variants in the MHC locus, appearing to explain the associations<sup>11,12</sup>. In contrast, in schizophrenia, classical HLA alleles



**Figure 2 | Haplotypes formed by *C4* structures and SNPs.** SNP haplotype(s) on which common *C4* structures were present. Each thin horizontal line represents the series of SNP alleles (haplotype) along a 250 kilobase (kb) chromosomal segment. Each column represents a SNP; grey and black indicate which allele is present on each haplotype. The SNP

haplotypes are grouped into 13 sets of haplotypes associating with each of the four most common *C4* structures. Three *C4* structures (AL-BS, AL-BL, and AL-AL) each segregated on multiple SNP haplotypes (numbered at right).

associated with schizophrenia less strongly than other genetic variants in the MHC region did (Extended Data Fig. 6). We further considered the strongest schizophrenia associations to classical *HLA* alleles at distinct loci (involving *HLA-B\*0801*, *HLA-DRB1\*0301*, and *HLA-DQB1\*02*); conditional analysis indicated that each could be explained by LD to the stronger signals at *C4* and rs13194504 (Extended Data Fig. 7).

If each *C4* allele affects schizophrenia risk via its effect on *C4A* expression, then this relationship should be visible across specific *C4* alleles. We measured schizophrenia risk levels for the common *C4* structural alleles (BS, AL-BS, AL-BL, and AL-AL); these alleles showed relative risks ranging from 1.00 to 1.27 (Fig. 5a). We also estimated (from the post-mortem brain samples) the *C4A* expression levels generated by these four alleles (Fig. 5b). Schizophrenia risk and *C4A* expression levels yielded the same ordering of the *C4* allelic series (Fig. 5a, b).

We sought an even more stringent test. If this allelic series of relationships with schizophrenia risk (Fig. 5a) arises from *C4* locus structure—rather than from other genetic variation in the MHC locus—then a given *C4* structure should exhibit the same schizophrenia risk regardless of the MHC haplotype on which it appears. We measured the schizophrenia association of all 13 common combinations of *C4* structure and MHC SNP haplotype (Fig. 5c). Across this allelic series, each *C4* allele exhibited a characteristic level of schizophrenia risk, regardless of the haplotype on which it was present (Fig. 5c).

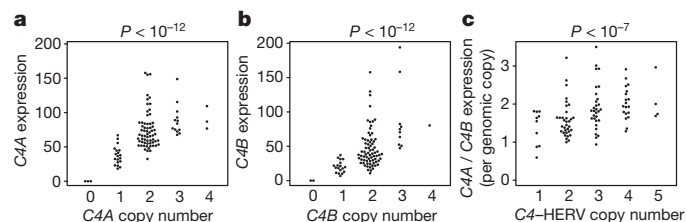
### *C4A* RNA expression in schizophrenia

These genetic findings (Fig. 5a, c) predict that *C4A* expression might be elevated in brain tissue from schizophrenia patients. We measured *C4A* RNA expression levels in brain tissue from 35 schizophrenia patients and 70 individuals without schizophrenia. The median expression of *C4A* in brain tissues from schizophrenia patients was 1.4-fold greater ( $P = 2 \times 10^{-5}$  by Mann-Whitney *U*-test; Fig. 5d) and was elevated in each of the five brain regions assayed (Extended Data Fig. 8). This relationship did not meaningfully change in analyses adjusted for age or post-mortem interval. The relationship remained significant after correcting for the higher average *C4A* copy number among the brain donors affected with schizophrenia (1.3-fold greater,  $P = 0.002$ ). Some earlier studies have also reported elevated levels of complement proteins in serum of schizophrenia patients<sup>27,28</sup>.

### *C4* in the central nervous system

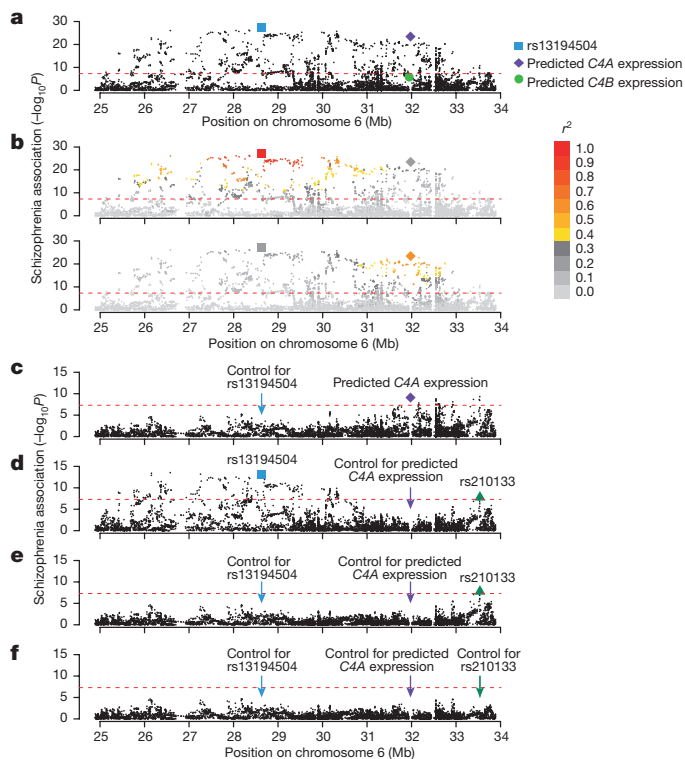
*C4* is a critical component of the classical complement cascade, an innate immune system pathway that rapidly recognizes and eliminates pathogens and cellular debris. In the brain, other genes in the classical complement cascade have been implicated in the elimination or ‘pruning’ of synapses<sup>29–31</sup>.

To evaluate the distribution of *C4* in human brain, we performed immunohistochemistry on sections of the prefrontal cortex and hippocampus. We observed *C4*<sup>+</sup> cells in the grey and white matter, with the greatest number of *C4*<sup>+</sup> cells detected in the hippocampus. Co-staining with cell-type-specific markers revealed *C4* in subsets of NeuN<sup>+</sup> neurons (Fig. 6a; antibody specificity further evaluated in



**Figure 3 | Brain RNA expression of *C4A* and *C4B* in relation to copy numbers of *C4A*, *C4B*, and the *C4-HERV*.** a, b, mRNA expression of *C4A* (a) and *C4B* (b) was measured (by ddPCR) in brain tissue from 244 individuals. The copy numbers of *C4A*, *C4B*, and the *C4-HERV* were measured (by ddPCR) in genomic DNA from the brain donors. The results were consistent across 8 panels of brain tissue representing 5 brain regions and 3 distinct sets of donors (one set shown here, with data from 101 individuals; all panels in Extended Data Fig. 4; a few outlier points are beyond the range of these plots but are shown in Extended Data Fig. 4). *P* values were obtained by a Spearman rank correlation test. In c, expression of *C4A* (per genomic copy) is normalized to expression of *C4B* (per genomic copy) to control for *trans*-acting influences shared by *C4A* and *C4B*.



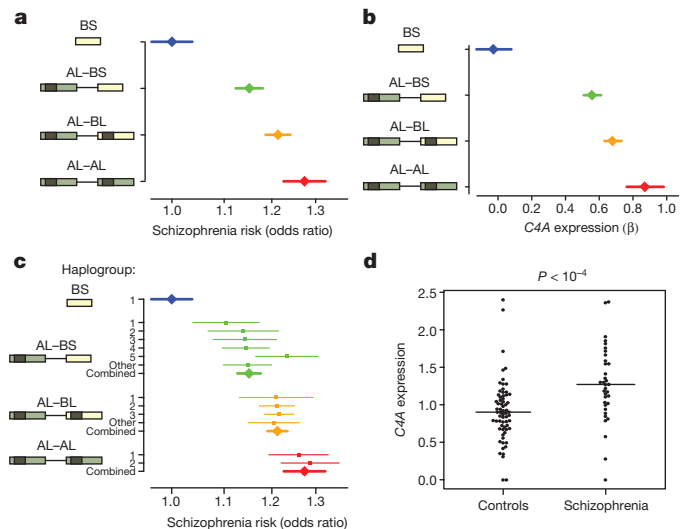


**Figure 4 | Association of schizophrenia to *C4* and the extended MHC locus.** Association of schizophrenia to 7,751 SNPs across the MHC locus and to genetically predicted expression levels of *C4A* and *C4B* in the brain (represented in the genomic location of the *C4* gene). The data shown are based on analysis of 28,799 schizophrenia cases and 35,986 controls of European ancestry from the Psychiatric Genomics Consortium. The height of each point represents the statistical strength ( $-\log_{10}(P)$ ) of association with schizophrenia. **a, b**, Association of schizophrenia to SNPs in the MHC locus and to genetically predicted expression of *C4A* and *C4B*. In **b**, genetic variants are coloured by their levels of correlation to rs13194504 (upper panel) or by their levels of correlation to genetically predicted brain *C4A* expression levels (lower panel). **c–f**, Conditional association analysis. The red dashed line indicates the statistical threshold for genome-wide significance ( $P = 5 \times 10^{-8}$ ). See also Extended Data Figs 5–7 for detailed association analyses involving *C4* locus structures and *HLA* alleles.

Extended Data Fig. 9a) and a subset of astrocytes. Much of the *C4* immunoreactivity was punctate (Fig. 6b), co-localizing with synaptic puncta identified by co-immunostaining for the pre- and postsynaptic markers VGLUT1/2 (also known as SLC17A7 and SLC17A6, respectively) and PSD-95 (also known as DLG4) (Fig. 6b). These results suggest that *C4* is produced by, or deposited on, neurons and synapses.

To further characterize neuronal *C4*, we cultured human primary cortical neurons and evaluated *C4* expression, localization, and secretion. Neurons expressed *C4* mRNA and secreted *C4* protein (Extended Data Fig. 9c). Neurons exhibited *C4*-immunoreactive puncta along their processes and cell bodies (Fig. 6c, d; antibody specificity further evaluated in Extended Data Fig. 9b). About 75% of *C4* immunoreactivity localized to neuronal processes (Fig. 6c); of the *C4* in neuronal processes, approximately 65% was observed in dendrites (MAP2<sup>+</sup>, neurofilament<sup>+</sup> processes) and 35% in axons (MAP2<sup>−</sup>, neurofilament<sup>+</sup> processes). Punctate *C4* immunoreactivity was observed at 48% of structural synapses as defined by co-localized synaptotagmin and PSD-95 (Fig. 6d).

The association of increased *C4* with schizophrenia (Figs 4 and 5), the presence of *C4* at synapses (Fig. 6b, d), the involvement of other complement proteins in synapse elimination<sup>29–31</sup>, and earlier reports of decreased synapse numbers in schizophrenia patients<sup>3–5</sup>, together suggested that *C4* might work with other components of the classical complement cascade to promote synaptic pruning. To test this hypothesis,



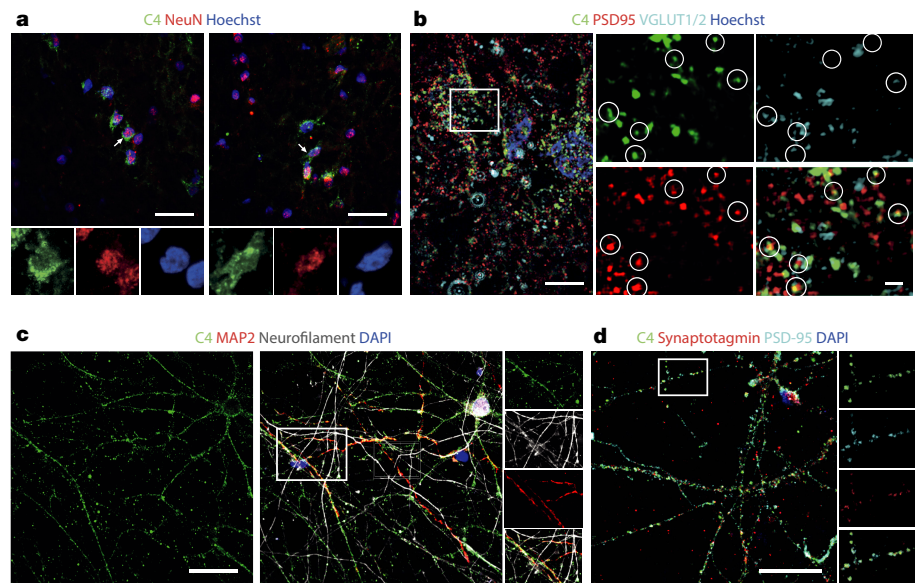
**Figure 5 | *C4* structures, *C4A* expression, and schizophrenia risk.** **a**, Schizophrenia risk associated with four common structural forms of *C4* in analysis of 28,799 schizophrenia cases and 35,986 controls. **b**, Brain *C4A* RNA expression levels associated with four common structural forms of *C4*.  $\beta$  was calculated from fitting *C4A* RNA expression (in brain tissue) to the number of chromosomes (0, 1, or 2) carrying each *C4* structure (across 120 individuals sampled). **c**, Schizophrenia risk associated with 13 combinations of *C4* structural allele and MHC SNP haplotype. The numbers on the y axis adjacent to the *C4* structures indicate the ‘haplogroup’, the MHC SNP haplotype background on which the *C4* structure segregates, and correspond to Fig. 2. Statistical tests of heterogeneity yielded  $P = 0.55$  for AL–AL alleles;  $P = 0.93$  for AL–BL alleles;  $P = 0.06$  for AL–BS alleles; and  $P = 5.7 \times 10^{-5}$  across the overall allelic series. **d**, Expression levels of *C4A* RNA were directly measured (by RT-ddPCR) in post-mortem brain samples from 35 schizophrenia patients and 70 individuals not affected with schizophrenia. Measurements for all five brain regions analysed exhibited the same relationship (Extended Data Fig. 8). Horizontal lines show the median value for each group.  $P$  values were derived by a (non-parametric) one-sided Mann–Whitney test. Error bars shown in **a–c** represent 95% confidence intervals around the effect size estimate.

we moved to a mouse model. *C4A* and *C4B* appear to have functionally specialized outside the rodent lineage, but the mouse genome contains a *C4* gene that shares features with both *C4A* and *C4B* (Extended Data Fig. 10a, b). Impairments in schizophrenia tend to affect higher cognitive functions and recently expanded brain regions for which analogies in mice are uncertain<sup>32</sup>. However, waves of postnatal synapse elimination occur in many brain regions, and strong experimental models have been established in several mammalian visual systems. In these systems, synaptic projections from retinal ganglion cells (RGCs) onto thalamic relay neurons within the dorsal lateral geniculate nucleus (dLGN) of the visual thalamus undergo activity-dependent synaptic refinement<sup>29–31,33–35</sup>. We found that *C4* RNA was expressed in the LGN and in RGCs purified from the retina during the period of synaptic pruning (Extended Data Fig. 10c).

In the immune system, *C4* promotes C3 activation, allowing C3 to covalently attach onto its targets and promote their engulfment by phagocytic cells. In the developing mouse brain, C3 targets subsets of synapses and is required for synapse elimination by microglia, the principal CNS cells expressing receptors for complement<sup>29,30</sup>. We found that in mice deficient in *C4* (ref. 36), C3 immunostaining in the dLGN was greatly reduced compared to wild-type littermates (Fig. 7a, b), with fewer synaptic inputs being C3-positive in the absence of *C4* (Fig. 7c). These data demonstrate a role for *C4* in complement deposition on synaptic inputs.

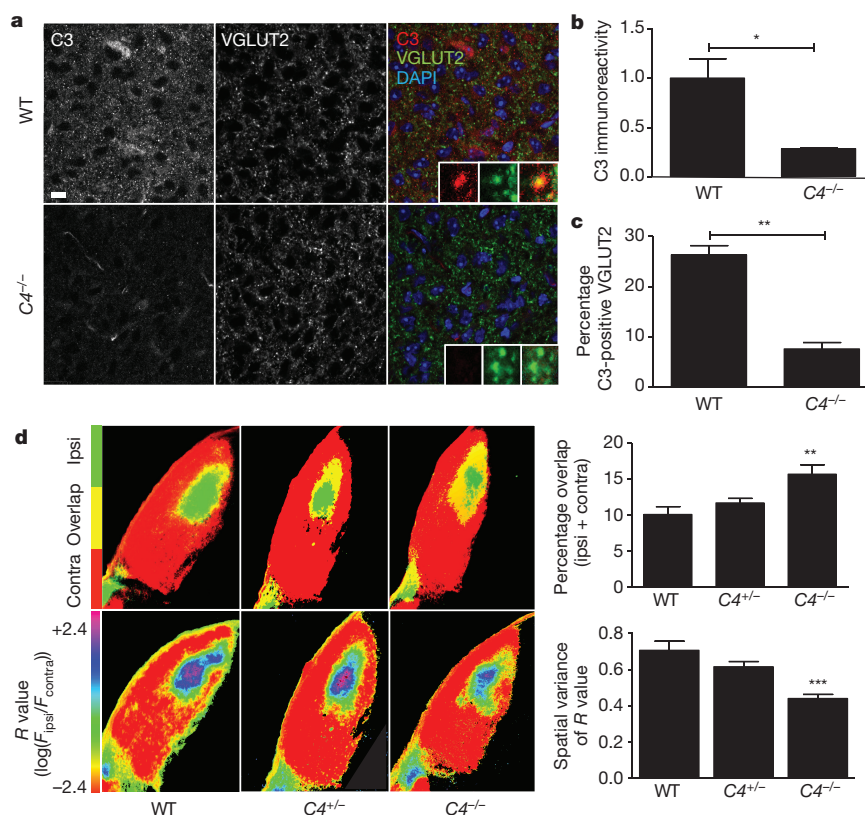
We then evaluated whether mice deficient in *C4* had defects in synaptic remodelling, as has been described for C3-deficient mice<sup>29</sup>. Mice lacking functional *C4* exhibited greater overlap between RGC inputs from the two eyes ( $P < 0.001$ ) than wild-type littermate controls,

**Figure 6 | C4 protein at neuronal cell bodies, processes and synapses.** **a**, C4 protein localization in human brain tissue. Two representative confocal images (drawn from immunohistochemistry performed on samples from five individuals with schizophrenia and two unaffected individuals) within the hippocampal formation demonstrate localization of C4 in a subset of NeuN<sup>+</sup> neurons. **b**, High-resolution structured illumination microscopy (SIM) imaging of tissue in the hippocampal formation reveals colocalization of C4 with the presynaptic terminal markers VGLUT1/2 and the postsynaptic marker PSD-95. **c**, Confocal images of primary human cortical neurons show colocalization of C4, MAP2, and neurofilament along neuronal processes. **d**, Confocal image of primary cortical neurons stained for C4, presynaptic marker synaptotagmin, and postsynaptic marker PSD-95. Scale bars, 25  $\mu$ m (**a**, **c**, and **d**); 5  $\mu$ m (**b**, left); and 1  $\mu$ m (**b**, right). Extended Data Fig. 9 contains additional data on antibody specificity.



suggesting reduced synaptic pruning (Fig. 7d, Extended Data Fig. 10d, e, and Supplementary Methods). The degree of deficit in  $C4^{-/-}$  mice was similar to that previously reported for  $C1q^{-/-}$  and

$C3^{-/-}$  mice<sup>29,31</sup>. Heterozygous  $C4^{+/-}$  mice, with one wild-type copy of C4, had an intermediate phenotype (Fig. 7d). These data provide direct evidence that C4 mediates synaptic refinement in the developing brain.



**Figure 7 | C4 in retinogeniculate synaptic refinement.** **a**, Representative confocal images of immunohistochemistry for C3 in the P5 dLGN showed reduced C3 deposition in the dLGN of  $C4^{-/-}$  mice compared to wild-type (WT) littermates. **b**, Quantification confirmed reduced C3 immunoreactivity in the dLGN ( $n = 3$  mice per group,  $P < 0.05$ ,  $t$ -test; y axis: mean fluorescence intensity, normalized to wild type). **c**, Co-localization analysis revealed a reduction in the fraction of VGLUT2<sup>+</sup> puncta that were C3<sup>+</sup> in  $C4$ -deficient mice relative to their WT littermates ( $n = 3$  mice per group,  $P = 0.0011$ , two-sided  $t$ -test). **d**, Synaptic refinement in mice with 0, 1, or 2 copies of C4. These images represent the segregation of ipsilateral and contralateral RGC projections to the dLGN; two analysis methods were used. Top, projections from

the ipsilateral (green) and contralateral (red) eyes show minimal overlap (yellow) in wild-type mice. The overlapping area is significantly increased in  $C4^{-/-}$  mice ( $n = 6$  mice per group,  $P < 0.01$ , ANOVA with Bonferroni post-hoc tests). Bottom, threshold-independent analysis using the  $R$  value<sup>50</sup> ( $R = \log_{10}(F_{\text{ipsi}}/F_{\text{contra}})$ ). Pixels are pseudocoloured with an  $R$  value heat map (red indicates areas having only contralateral inputs; purple, only ipsilateral inputs). Compared to their wild-type littermates,  $C4$ -deficient mice exhibited lower  $R$  value variance, indicating defects in synaptic refinement ( $n = 6$  mice per group,  $P < 0.001$ , ANOVA with Bonferroni post-hoc tests). Control experiments analysing total dLGN size, dLGN area receiving ipsilateral input, and number of RGCs are shown in Extended Data Fig. 10f–h, respectively. Error bars in **b–d** represent s.e.m.



## Discussion

We developed ways to analyse a complex form of genome structural variation (Figs 1 and 2) and discovered that schizophrenia's association with variation in the MHC locus involves many common, structurally distinct *C4* alleles that affect expression of *C4A* and *C4B* in the brain; each allele associated with schizophrenia risk in proportion to its effect on *C4A* expression (Figs 3–5). We found that *C4* is expressed by neurons, localized to dendrites, axons, and synapses, and is secreted (Fig. 6). In mice, *C4* promoted synapse elimination during the developmentally timed maturation of a neuronal circuit (Fig. 7).

In humans, adolescence and early adulthood bring extensive elimination of synapses in distributed association regions of cerebral cortex, such as the prefrontal cortex, that have greatly expanded in recent human evolution and appear to become impaired in schizophrenia<sup>37–40</sup>. Synapse elimination in human association cortex appears to continue from adolescence into the third decade of life<sup>39</sup>. This late phase of cortical maturation, which may distinguish humans even from some other primates<sup>37</sup>, corresponds to the period during which schizophrenia most often becomes clinically apparent and patients' cognitive function declines, a temporal correspondence that others have also noted<sup>41</sup>.

The principal pathological findings in brains of those affected with schizophrenia involve loss of cortical grey matter without cell death: affected individuals exhibit abnormal cortical thinning<sup>1,2</sup> and reduced numbers of synaptic structures on cortical pyramidal neurons<sup>3–5</sup>. In the brain, complement receptors are expressed primarily by microglia, the phagocytic immune cells of the central nervous system. The possibility that neuron–microglia interactions via the complement cascade contribute to schizophrenia pathogenesis—for example, that schizophrenia arises or intensifies from excessive or inappropriate synaptic pruning by microglia during adolescence and early adulthood—would offer a potential mechanism for these longstanding observations about age of onset and synapse loss. Many other genetic findings in schizophrenia involve genes that encode synaptic proteins<sup>6,42–44</sup>. Diverse synaptic abnormalities could in principle interact with the complement system and other pathways<sup>45,46</sup> to cause excessive stimulation of microglia and elimination of synapses during adolescence and early adulthood.

The two human *C4* genes (*C4A* and *C4B*) exhibited distinct relationships with schizophrenia risk, with increased risk associating most strongly with variation that increases expression of *C4A*. Human *C4A* and *C4B* proteins, whose functional specialization appears to be evolutionarily recent (Extended Data Fig. 10a), show striking biochemical differences: *C4A* more readily forms amide bonds with proteins, while *C4B* favours binding to carbohydrate surfaces<sup>19,20</sup>, differences with an established basis in *C4* protein sequence and structure<sup>47,48</sup>. An intriguing possibility is that *C4A* and *C4B* differ in affinity for an unknown binding site at synapses.

To date, few genome-wide association study (GWAS) associations have been explained by specific functional alleles. An unexpected finding at *C4* involves the large number of common, functionally distinct forms of the same locus that appear to contribute to schizophrenia risk. The human genome contains hundreds of other genes with complex, multi-allelic forms of structural variation<sup>49</sup>. It will be important to learn the extent to which such variation contributes to brain diseases and to all human phenotypes.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 19 May; accepted 18 December 2015.**

**Published online 27 January 2016.**

1. Cannon, T. D. *et al.* Cortex mapping reveals regionally specific patterns of genetic and disease-specific gray-matter deficits in twins discordant for schizophrenia. *Proc. Natl Acad. Sci. USA* **99**, 3228–3233 (2002).
2. Cannon, T. D. *et al.* Progressive reduction in cortical thickness as psychosis develops: a multisite longitudinal neuroimaging study of youth at elevated clinical risk. *Biol. Psychiatry* **77**, 147–157 (2015).
3. Garey, L. J. *et al.* Reduced dendritic spine density on cerebral cortical pyramidal neurons in schizophrenia. *J. Neurol. Neurosurg. Psychiatry* **65**, 446–453 (1998).

4. Glantz, L. A. & Lewis, D. A. Decreased dendritic spine density on prefrontal cortical pyramidal neurons in schizophrenia. *Arch. Gen. Psychiatry* **57**, 65–73 (2000).
5. Glausier, J. R. & Lewis, D. A. Dendritic spine pathology in schizophrenia. *Neuroscience* **251**, 90–107 (2013).
6. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
7. Shi, J. *et al.* Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* **460**, 753–757 (2009).
8. Stefansson, H. *et al.* Common variants conferring risk of schizophrenia. *Nature* **460**, 744–747 (2009).
9. International Schizophrenia Consortium *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
10. Schizophrenia Psychiatric Genome-Wide Association Study Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nature Genet.* **43**, 969–976 (2011).
11. Howson, J. M., Walker, N. M., Clayton, D. & Todd, J. A. Confirmation of HLA class II independent type 1 diabetes associations in the major histocompatibility complex including HLA-B and HLA-A. *Diabetes Obes. Metab.* **11** (Suppl 1), 31–45 (2009).
12. Raychaudhuri, S. *et al.* Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nature Genet.* **44**, 291–296 (2012).
13. Escudero-Esparza, A., Kalchishkova, N., Kurbasic, E., Jiang, W. G. & Blom, A. M. The novel complement inhibitor human CUB and Sushi multiple domains 1 (CSMD1) protein promotes factor I-mediated degradation of C4b and C3b and inhibits the membrane attack complex assembly. *FASEB J.* **27**, 5083–5093 (2013).
14. Carroll, M. C., Campbell, R. D., Bentley, D. R. & Porter, R. R. A molecular map of the human major histocompatibility complex class III region linking complement genes *C4*, *C2* and factor *B*. *Nature* **307**, 237–241 (1984).
15. Carroll, M. C., Belt, T., Palsdottir, A. & Porter, R. R. Structure and organization of the *C4* genes. *Phil. Trans. R. Soc. Lond. B* **306**, 379–388 (1984).
16. Dangel, A. W. *et al.* The dichotomous size variation of human complement *C4* genes is mediated by a novel family of endogenous retroviruses, which also establishes species-specific genomic patterns among Old World primates. *Immunogenetics* **40**, 425–436 (1994).
17. Horton, R. *et al.* Variation analysis and gene annotation of eight MHC haplotypes: the MHC Haplotype Project. *Immunogenetics* **60**, 1–18 (2008).
18. Bánlák, Z., Doleschall, M., Rajczy, K., Fust, G. & Szilágyi, A. Fine-tuned characterization of RCCX copy number variants and their relationship with extended MHC haplotypes. *Genes Immun.* **13**, 530–535 (2012).
19. Law, S. K., Dodds, A. W. & Porter, R. R. A comparison of the properties of two classes, *C4A* and *C4B*, of the human complement component *C4*. *EMBO J.* **3**, 1819–1823 (1984).
20. Isenman, D. E. & Young, J. R. The molecular basis for the difference in immune hemolysis activity of the Chido and Rodgers isotypes of human complement component *C4*. *J. Immunol.* **132**, 3019–3027 (1984).
21. Illarionova, A. E., Vinogradova, T. V. & Sverdlov, E. D. Only those genes of the KIAA1245 gene subfamily that contain HERV(K) LTRs in their introns are transcriptionally active. *Virology* **358**, 39–47 (2007).
22. Nakamura, A., Okazaki, Y., Sugimoto, J., Oda, T. & Jinno, Y. Human endogenous retroviruses with transcriptional potential in the brain. *J. Hum. Genet.* **48**, 575–581 (2003).
23. Suntsova, M. *et al.* Human-specific endogenous retroviral insert serves as an enhancer for the schizophrenia-linked gene *PRODH*. *Proc. Natl Acad. Sci. USA* **110**, 19472–19477 (2013).
24. Yang, Y. *et al.* Diversity in intrinsic strengths of the human complement system: serum *C4* protein concentrations correlate with *C4* gene size and polygenic variations, hemolytic activities, and body mass index. *J. Immunol.* **171**, 2734–2745 (2003).
25. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
26. Iossifov, I. *et al.* The contribution of *de novo* coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).
27. Mayilyan, K. R., Arnold, J. N., Presanis, J. S., Soghoyan, A. F. & Sim, R. B. Increased complement classical and mannan-binding lectin pathway activities in schizophrenia. *Neurosci. Lett.* **404**, 336–341 (2006).
28. Hakobyan, S., Boyajyan, A. & Sim, R. B. Classical pathway complement activity in schizophrenia. *Neurosci. Lett.* **374**, 35–37 (2005).
29. Stevens, B. *et al.* The classical complement cascade mediates CNS synapse elimination. *Cell* **131**, 1164–1178 (2007).
30. Schafer, D. P. *et al.* Microglia sculpt postnatal neural circuits in an activity and complement-dependent manner. *Neuron* **74**, 691–705 (2012).
31. Bialas, A. R. & Stevens, B. TGF- $\beta$  signaling regulates neuronal C1q expression and developmental synaptic refinement. *Nature Neurosci.* **16**, 1773–1782 (2013).
32. Kaiser, T. & Feng, G. Modeling psychiatric disorders for developing effective treatments. *Nature Med.* **21**, 979–988 (2015).
33. Shatz, C. J. & Kirkwood, P. A. Prenatal development of functional connections in the cat's retinogeniculate pathway. *J. Neurosci.* **4**, 1378–1397 (1984).
34. Sretavan, D. W. & Shatz, C. J. Prenatal development of retinal ganglion cell axons: segregation into eye-specific layers within the cat's lateral geniculate nucleus. *J. Neurosci.* **6**, 234–251 (1986).



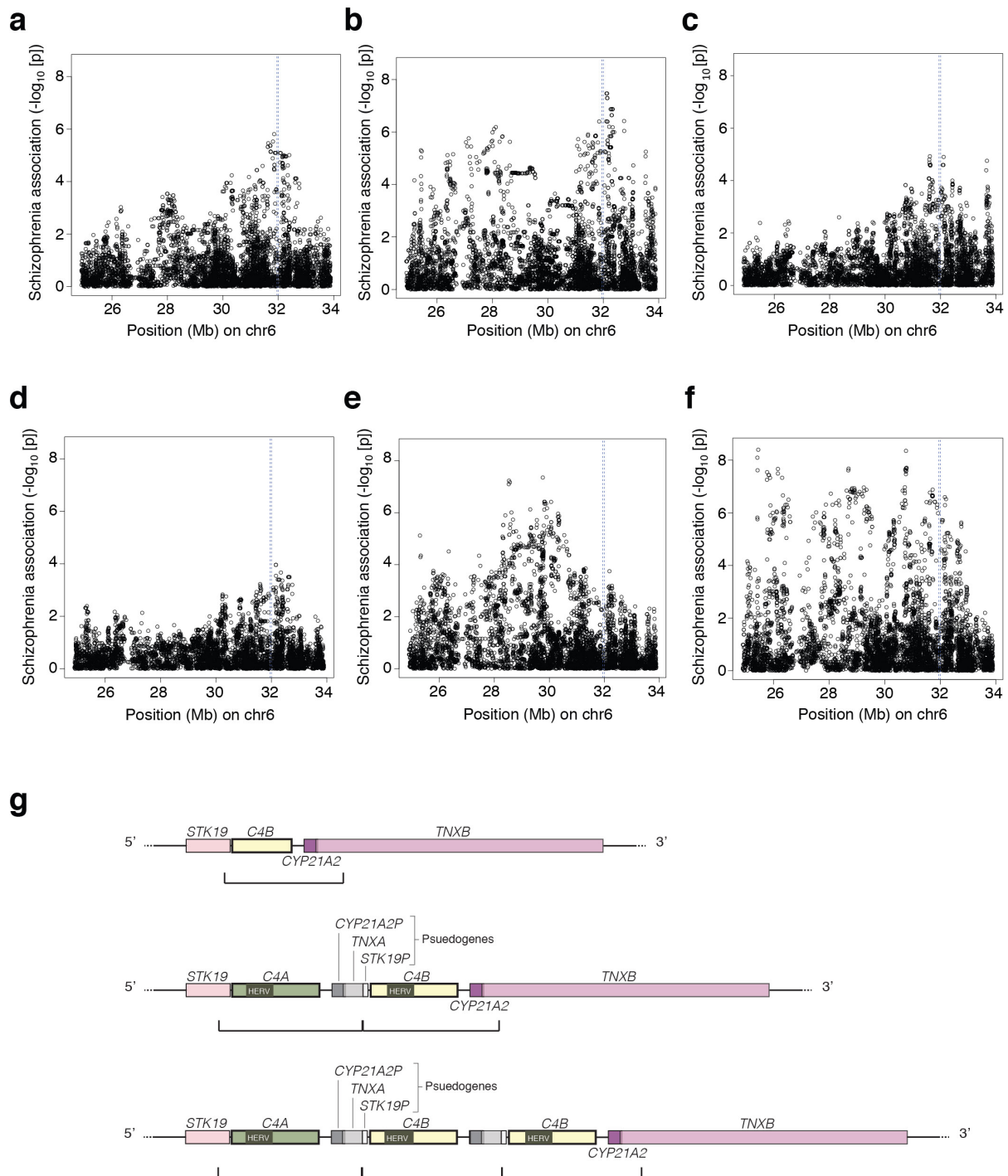
35. Chen, C. & Regehr, W. G. Developmental remodeling of the retinogeniculate synapse. *Neuron* **28**, 955–966 (2000).
36. Fischer, M. B. *et al.* Regulation of the B cell response to T-dependent antigens by classical pathway complement. *J. Immunol.* **157**, 549–556 (1996).
37. Huttenlocher, P. R. & Dabholkar, A. S. Regional differences in synaptogenesis in human cerebral cortex. *J. Comp. Neurol.* **387**, 167–178 (1997).
38. Huttenlocher, P. R. Synaptic density in human frontal cortex—developmental changes and effects of aging. *Brain Res.* **163**, 195–205 (1979).
39. Petanjek, Z. *et al.* Extraordinary neoteny of synaptic spines in the human prefrontal cortex. *Proc. Natl Acad. Sci. USA* **108**, 13281–13286 (2011).
40. Buckner, R. L. & Krienen, F. M. The evolution of distributed association networks in the human brain. *Trends Cogn. Sci.* **17**, 648–665 (2013).
41. Feinberg, I. Schizophrenia: caused by a fault in programmed synaptic elimination during adolescence? *J. Psychiatr. Res.* **17**, 319–334 (1982–1983).
42. Kirov, G. *et al.* *De novo* CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol. Psychiatry* **17**, 142–153 (2012).
43. Fromer, M. *et al.* *De novo* mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179–184 (2014).
44. Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185–190 (2014).
45. Datwani, A. *et al.* Classical MHCI molecules regulate retinogeniculate refinement and limit ocular dominance plasticity. *Neuron* **64**, 463–470 (2009).
46. Lee, H. *et al.* Synapse elimination and learning rules co-regulated by MHC class I H2-Db. *Nature* **509**, 195–200 (2014).
47. van den Elsen, J. M. *et al.* X-ray crystal structure of the C4d fragment of human complement component C4. *J. Mol. Biol.* **322**, 1103–1115 (2002).
48. Dodds, A. W., Ren, X. D., Willis, A. C. & Law, S. K. The reaction mechanism of the internal thioester in the human complement component C4. *Nature* **379**, 177–179 (1996).
49. Handsaker, R. E. *et al.* Large multiallelic copy number variations in humans. *Nature Genet.* **47**, 296–303 (2015).
50. Torborg, C. L. & Feller, M. B. Unbiased analysis of bulk axonal segregation patterns. *J. Neurosci. Methods* **135**, 17–26 (2004).
51. Fernando, M. M. *et al.* Assessment of complement C4 gene copy number using the paralog ratio test. *Hum. Mutat.* **31**, 866–874 (2010).
52. Rudduck, C., Beckman, L., Franzen, G., Jacobsson, L. & Lindstrom, L. Complement factor C4 in schizophrenia. *Hum. Hered.* **35**, 223–226 (1985).
53. Schroers, R. *et al.* Investigation of complement C4B deficiency in schizophrenia. *Hum. Hered.* **47**, 279–282 (1997).
54. Mayilyan, K. R., Dodds, A. W., Boyajyan, A. S., Soghoyan, A. F. & Sim, R. B. Complement C4B protein in schizophrenia. *World J. Biol. Psychiatry* **9**, 225–230 (2008).
55. Jia, X. *et al.* Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS ONE* **8**, e64683 (2013).
56. Nonaka, M., Nakayama, K., Yeul, Y. D. & Takahashi, M. Complete nucleotide and derived amino acid sequences of sex-limited protein (Slp), nonfunctional isotype of the fourth component of mouse complement (C4). *J. Immunol.* **136**, 2989–2993 (1986).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** The authors would like to remember the late T. Stanley with appreciation and express their gratitude for his support. We thank S. Hyman, E. Lander, C. Bargmann, and C. Patil for conversations about the project and comments on drafts of the manuscript; M. Webster for expert advice on immunohistochemistry; B. Browning for expert advice on imputation; the Stanley Medical Research Institute Brain Collection and the NHGRI Gene and Tissue Expression (GTEx) Project for access to RNA and tissue samples; C. Emba for assistance with experiments; and C. Usher for contributions to manuscript figures. This work was supported by R01 HG 006855 (to S.A.M.), by the Stanley Center for Psychiatric Research (to S.A.M. and B.S.), by U01 MH105641 (to S.A.M.), by R01 MH077139 (to the PGC), and by T32 GM007753 (to A.S. and M.B.).

**Author Contributions** S.A.M. and A.S. conceived the genetic studies. A.S. performed the laboratory experiments and computational analyses to understand the molecular and population genetics of the C4 locus (Figs 1 and 2). A.S., K.T., N.K., and V.V.D. analysed C4 expression variation in human brain (Figs 3 and 5b, d). G.G., R.E.H., and S.A.R. contributed to genetic analyses. A.S. and A.D. did the imputation and association analysis (Figs 4 and 5a, c). M.J.D. provided advice on the association analyses. Investigators in the Schizophrenia Working Group of the Psychiatric Genomics Consortium collected and phenotyped cohorts and contributed genotype data for analysis. B.S. and M.C.C. contributed expertise and reagents for experiments described in Fig. 6 and 7. H.d.R. and T.R.H. performed the C4 immunocytochemistry and immunohistochemistry experiments respectively, with advice from A.R.B. (Fig. 6). A.R.B. and J.P. analysed the role of C4 in synaptic refinement in the mouse visual system (Fig. 7). M.B. analysed C4 expression in mice. S.A.M. and A.S. wrote the manuscript with contributions from all authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.A.M. ([mccarroll@genetics.med.harvard.edu](mailto:mccarroll@genetics.med.harvard.edu)).

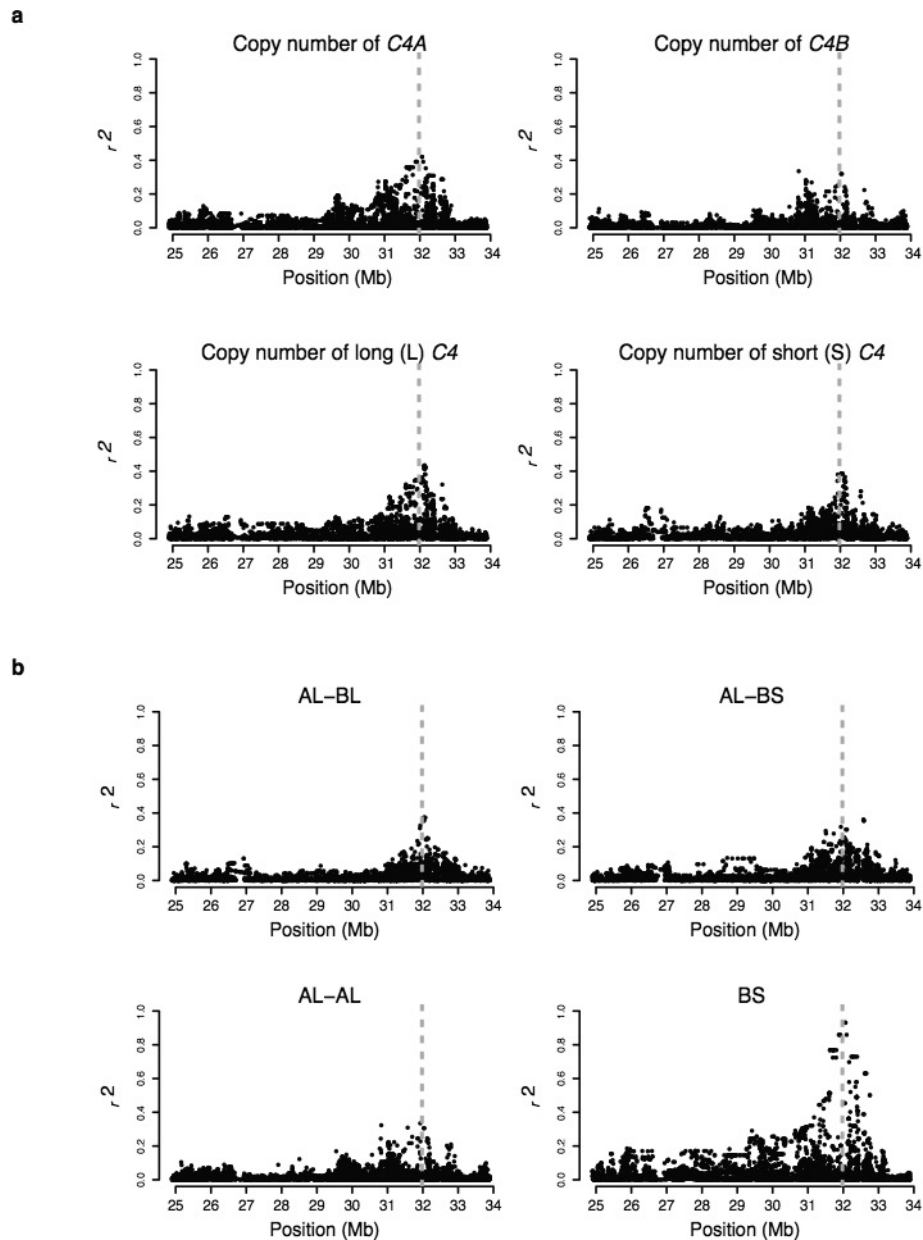


**Extended Data Figure 1 | Association of schizophrenia to common variants in the MHC locus in individual case-control cohorts, and schematic of the repeat module containing C4.** a–f, Data for several schizophrenia case-control cohorts that were genome-scanned before we began this work (a–d) exhibits peaks of association near chr6: 32 Mb (blue vertical line) on the human genome reference sequence (GRCh37/hg19). Note that association patterns vary from cohort to cohort, reflecting statistical sampling fluctuations and potentially fluctuations in allele frequencies of the (unknown) causal variants in different cohorts. Cohorts such as in b, e and f suggest the existence of effects at multiple loci within the MHC region. Even in the cohorts with simpler peaks (a, c, d), the pattern of association across the individual SNPs at chr6: 32 Mb does not correspond to the LD around any known variant. This motivated the focus in the current work on cryptic genetic influences in this region that could cause unconventional association signals that do not resemble the LD patterns of individual variants. g, A complex form of genome structural variation resides near chr6:

32 Mb. Shown here are three of the known alternative structural forms of this genomic region. The most prominent feature of this structural variation is the tandem duplication of a genomic segment that contains a C4 gene, 3' fragments of the STK19 and TNXB genes, and a pseudogenized copy of the CYP21A2 gene. This cassette is present in 1–3 copies on the three alleles depicted above; the boundaries below each haplotype demarcate the sequence that is duplicated. Haplotypes with multiple copies of this module (middle and bottom) contain multiple functional copies of C4, whereas the additional gene fragments or copies denoted STK19P, CYP21A2P, and TNXA are typically pseudogenized. Rare haplotypes with a gain or loss of intact CYP21A2 have also been observed<sup>18</sup>. Although C4A and C4B contain multiple sequence variants, they are defined based on the differences encoded by exon 26, which determine the relative affinities of C4A and C4B for distinct molecular targets<sup>19,20</sup> (Fig. 1). Many additional forms of this locus appear to have arisen by non-allelic homologous recombination and gene conversion (ref. 18 and Fig. 1).

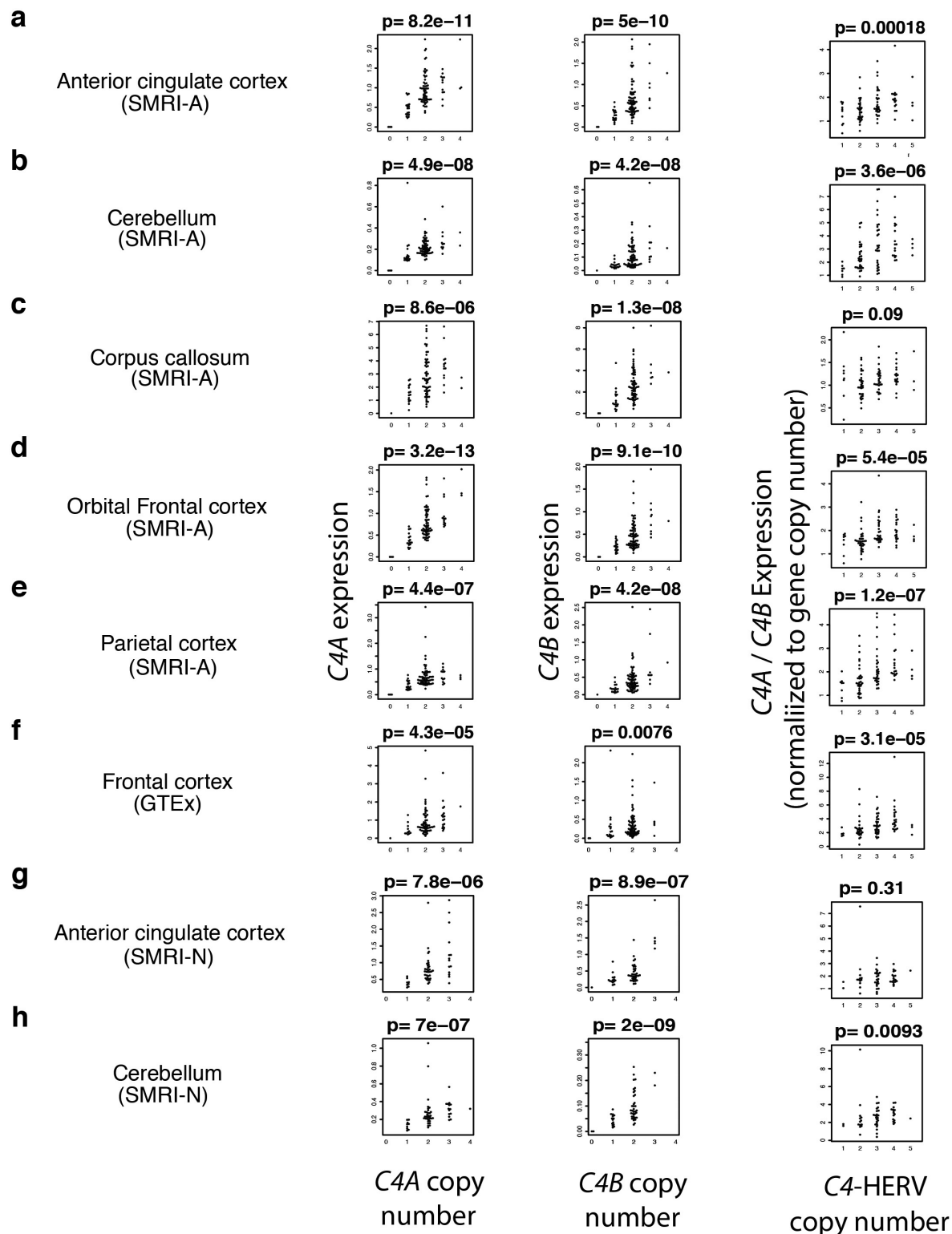






**Extended Data Figure 3 | Linkage disequilibrium relationships ( $r^2$ ) of MHC SNPs to forms of *C4* structural variation. a, b, Correlations of SNPs in the MHC locus with copy number of *C4* gene types (a) and larger-scale structural forms (haplotypes) (b) of the *C4* locus. Dashed,**

**vertical lines indicate the genomic location of the *C4* locus. *C4* structural forms show only partial correlation ( $r^2$ ) to the allelic states of nearby SNPs, reflecting the relationship shown in Fig. 2, in which a structural form of the *C4* locus often segregates on multiple different SNP haplotypes.**



**Extended Data Figure 4 | RNA expression of *C4A* and *C4B* in relation to copy number of *C4A*, *C4B*, and the *C4*-HERV (long form of *C4*), in eight panels of post-mortem brain tissue.** Copy number of *C4* structural features was measured by ddPCR; RNA expression levels were measured by RT-ddPCR. **a–e**, Data for tissues from the Stanley Medical Research Institute (SMRI) Array Consortium consisting of anterior cingulate cortex (**a**), cerebellum (**b**), corpus callosum (**c**), orbital frontal cortex (**d**), and parietal cortex (**e**). **f**, Data for the frontal cortex samples from the NHGRI Genes and Tissues Expression (GTEx) Project. **g, h**, Data for tissues from the SMRI Neuropathology Consortium (anterior cingulate cortex and cerebellum, respectively). These data were then used to inform (by linear

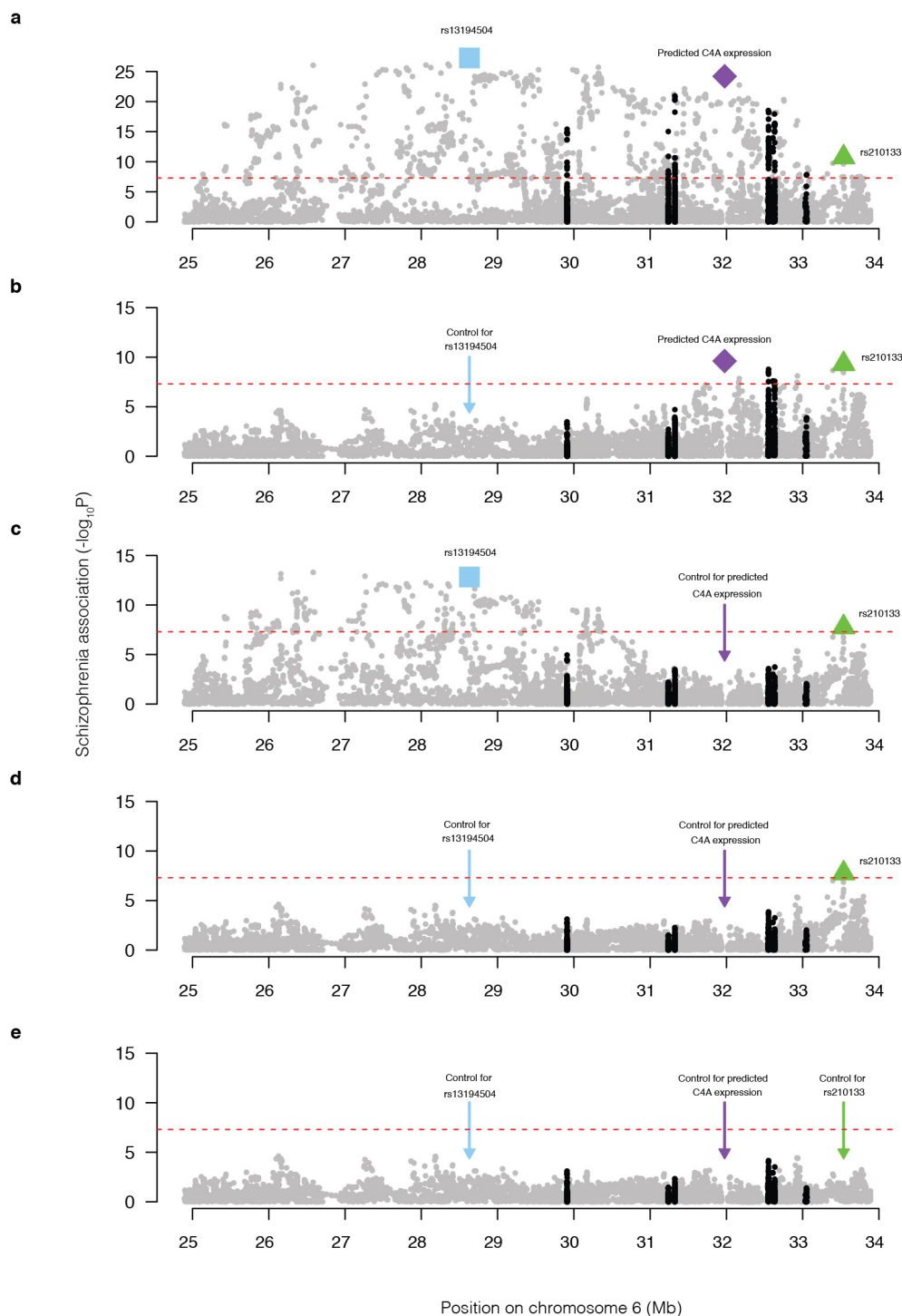
regression) the derivation of a linear model for predicting each individual's RNA expression of *C4A* and *C4B* as a function of the numbers of copies of AL, BL, AS, and BS. The derivation of this model, and the regression coefficients induced, are described in Supplementary Methods. In the rightmost plot of each panel, expression of *C4A* (per genomic copy) is normalized to expression of *C4B* (per genomic copy) to more specifically visualize the effect of the *C4*-HERV by controlling for genomic copy number and for any *trans*-acting influences shared by *C4A* and *C4B*; the inferred regression coefficients (Supplementary Methods) suggest that the observed effect is mostly due to increased expression of *C4A*.

	Association of SCZ with this variant (unconditioned analysis)			Association of SCZ with this variant (joint analysis with predicted <i>C4A</i> expression as a covariate)			Association of SCZ with genetically predicted <i>C4A</i> expression (joint analysis with this variant as a covariate)		
	p	β	SE	p	β	SE	p	β	SE
<b>Expression predictors (derived from <i>post mortem</i> brain RNA analysis)</b>									
<i>C4A</i> expression (genetic prediction from imputed <i>C4</i> locus structure)	3.60E-24	0.247	0.024	NA	NA	NA	NA	NA	NA
<i>C4B</i> expression (genetic prediction from imputed <i>C4</i> locus structure)	2.31E-07	-0.092	0.018	0.71	0.008	0.021	2.35E-18	0.253	0.029
<b>Copy number of <i>C4</i> structural features</b>									
<i>C4A</i> genes	6.31E-19	0.112	0.013	0.45	-0.023	0.030	6.54E-07	0.287	0.058
<i>C4B</i> genes	1.45E-04	-0.066	0.017	0.63	0.009	0.019	4.58E-21	0.252	0.027
Long <i>C4</i> genes (with HERV)	7.09E-23	0.084	0.009	0.17	0.029	0.021	4.89E-03	0.170	0.060
Short <i>C4</i> genes (no HERV)	3.94E-14	-0.085	0.011	0.32	-0.015	0.015	7.38E-12	0.225	0.033
Total <i>C4</i> copy number	9.41E-15	0.123	0.016	1.00	0.000	0.025	5.40E-11	0.247	0.038
<b>Specific <i>C4</i> locus structures (haplotypes of one or more <i>C4</i> genes)</b>									
BS	2.29E-19	-0.171	0.019	0.16	-0.045	0.032	1.10E-06	0.200	0.041
AL-BS	0.03	-0.027	0.013	0.62	0.007	0.013	3.39E-23	0.250	0.025
AL-BL	6.16E-07	0.058	0.012	0.41	0.011	0.013	6.34E-19	0.238	0.027
AL-AL	2.09E-06	0.093	0.020	0.56	-0.013	0.023	2.49E-19	0.256	0.028
<b>Combinations of <i>C4</i> locus structure and flanking SNP haplotype</b>									
AL-BS-4	0.21	-0.027	0.022	0.95	-0.002	0.022	7.99E-24	0.247	0.025
AL-BS-2	0.35	-0.031	0.033	0.80	-0.008	0.033	5.45E-24	0.247	0.024
AL-BS-3	0.34	-0.028	0.029	0.91	-0.003	0.030	5.69E-24	0.247	0.024
AL-BS-5	0.07	0.053	0.029	0.01	0.077	0.029	5.73E-25	0.252	0.024
AL-BS-1	0.04	-0.064	0.031	0.19	-0.040	0.031	1.32E-23	0.245	0.024
AL-BS-other	0.30	-0.023	0.022	0.92	0.002	0.023	6.19E-24	0.247	0.025
AL-BL-2	0.03	0.036	0.017	0.77	0.005	0.017	3.52E-23	0.246	0.025
AL-BL-3	2.15E-03	0.042	0.014	0.55	0.008	0.014	3.28E-22	0.243	0.025
AL-BL-1	0.35	0.032	0.034	0.92	0.003	0.035	5.55E-24	0.247	0.024
AL-BL-other	0.23	0.029	0.024	0.95	0.002	0.024	7.37E-24	0.247	0.025
AL-AL-1	0.01	0.074	0.028	0.38	-0.026	0.029	9.42E-23	0.255	0.026
AL-AL-2	3.33E-04	0.097	0.027	0.94	-0.002	0.029	2.31E-21	0.248	0.026
<b>Alternative hypotheses</b>									
<i>C4B</i> nulls (for whom total <i>C4B</i> copy number = 0) *	0.36	0.061	0.066						

**Extended Data Figure 5 | Detailed analysis of the association of schizophrenia to genetic variation at and around *C4*, in data from 28,799 schizophrenia cases and 35,986 controls.** (Psychiatric Genomics Consortium, ref. 6.) SCZ, schizophrenia; β, estimated effect size per copy of the genomic feature or allele indicated; SE, standard error. Detailed association analyses of *HLA* alleles are in Extended Data Figs 6 and 7. The single asterisk (\*) indicates that we specifically tested *C4B*-null status because a 1985 study<sup>52</sup> reported an analysis of 165 schizophrenia patients

and 330 controls in which rare *C4B*-null status associated with elevated risk of schizophrenia, though two subsequent studies<sup>53,54</sup> found no association of schizophrenia to *C4B*-null genotype. We sought to evaluate this using the large data set in this study, finding no association to *C4B*-null status. The double asterisk (\*\*) indicates total copy number of *C4* is also strongly correlated to copy number of the *CYP21A2P* pseudogene, which is present on duplicated copies of the sequence shown in Extended Data Fig. 1g.





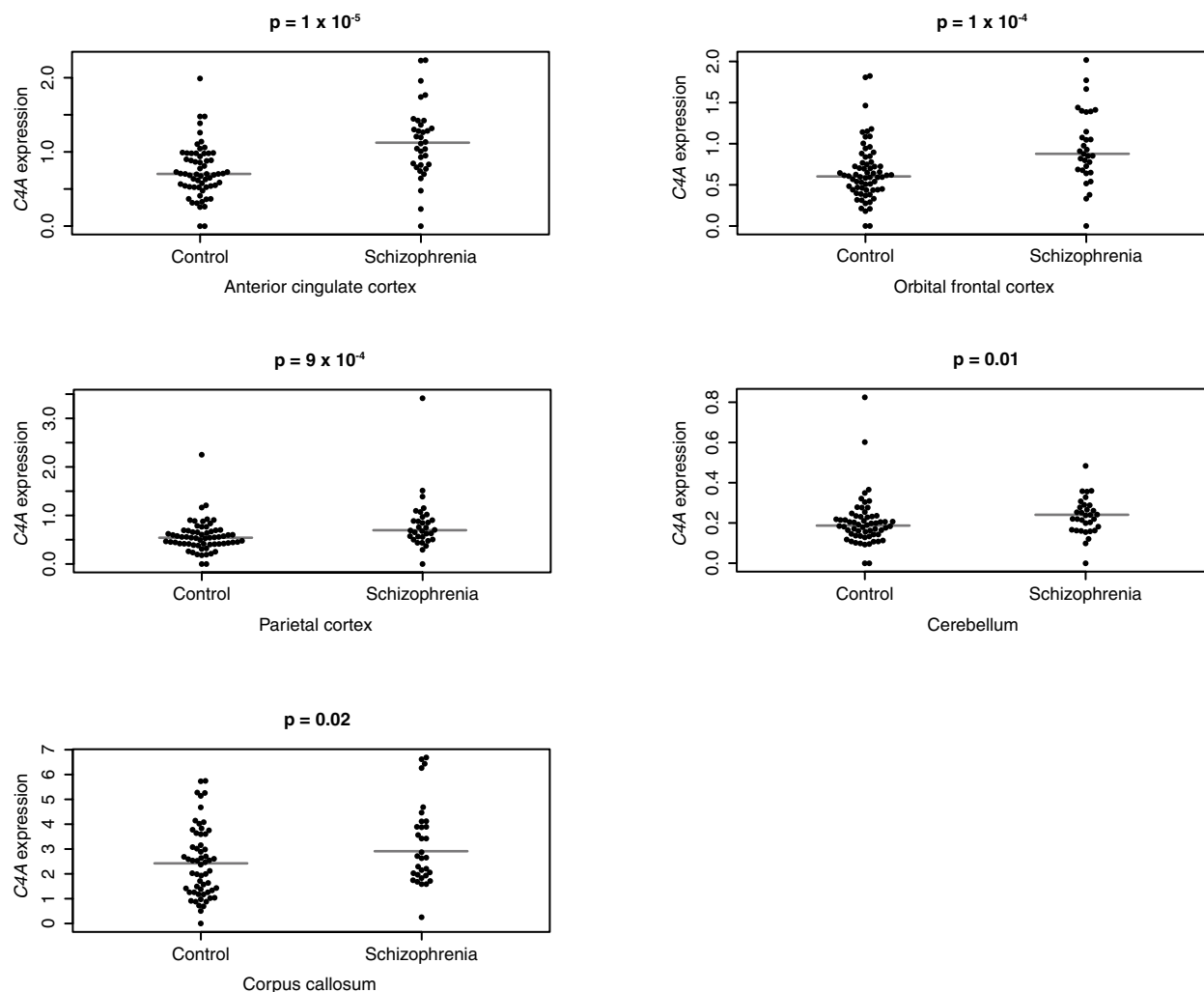
**Extended Data Figure 6 | Evaluation of the association of schizophrenia with *HLA* alleles and coding-sequence polymorphisms.** **a–e**, Associations to *HLA* alleles and coding-sequence polymorphisms are shown in black; to provide the context of levels of association to

nearby SNPs, associations to other SNPs are shown in grey. The series of conditional analyses shown in **b–e** parallels the analyses in Fig. 4. Further detail on the most strongly associating *HLA* alleles (including conditional association analysis) is provided in Extended Data Fig. 7.

Variants included in analysis	Association p-value				
	rs13194504	Genetically predicted <i>C4A</i> expression (C4Aexp)	<i>B*0801</i>	<i>DRB1*0301</i>	<i>DQB1*02</i>
rs13194504	5.50E-28				
C4Aexp		3.60E-24			
B*0801			1.20E-21		
DRB1*0301				6.40E-19	
DQB1*02					6.25E-17
rs13194504, C4Aexp	8.00E-14	7.80E-10			
rs13194504, B*0801	5.75E-11		2.00E-04		
rs13194504, DRB1*0301	2.89E-15			6.20E-06	
rs13194504, DQB1*02	9.60E-19				1.82E-07
rs13194504, C4Aexp, B*0801	6.94E-11	8.23E-07	0.96		
rs13194504, C4Aexp, DRB1*0301	1.39E-12	2.18E-05		0.41	
rs13194504, C4Aexp, DQB1*02	3.06E-13	1.03E-04			0.03

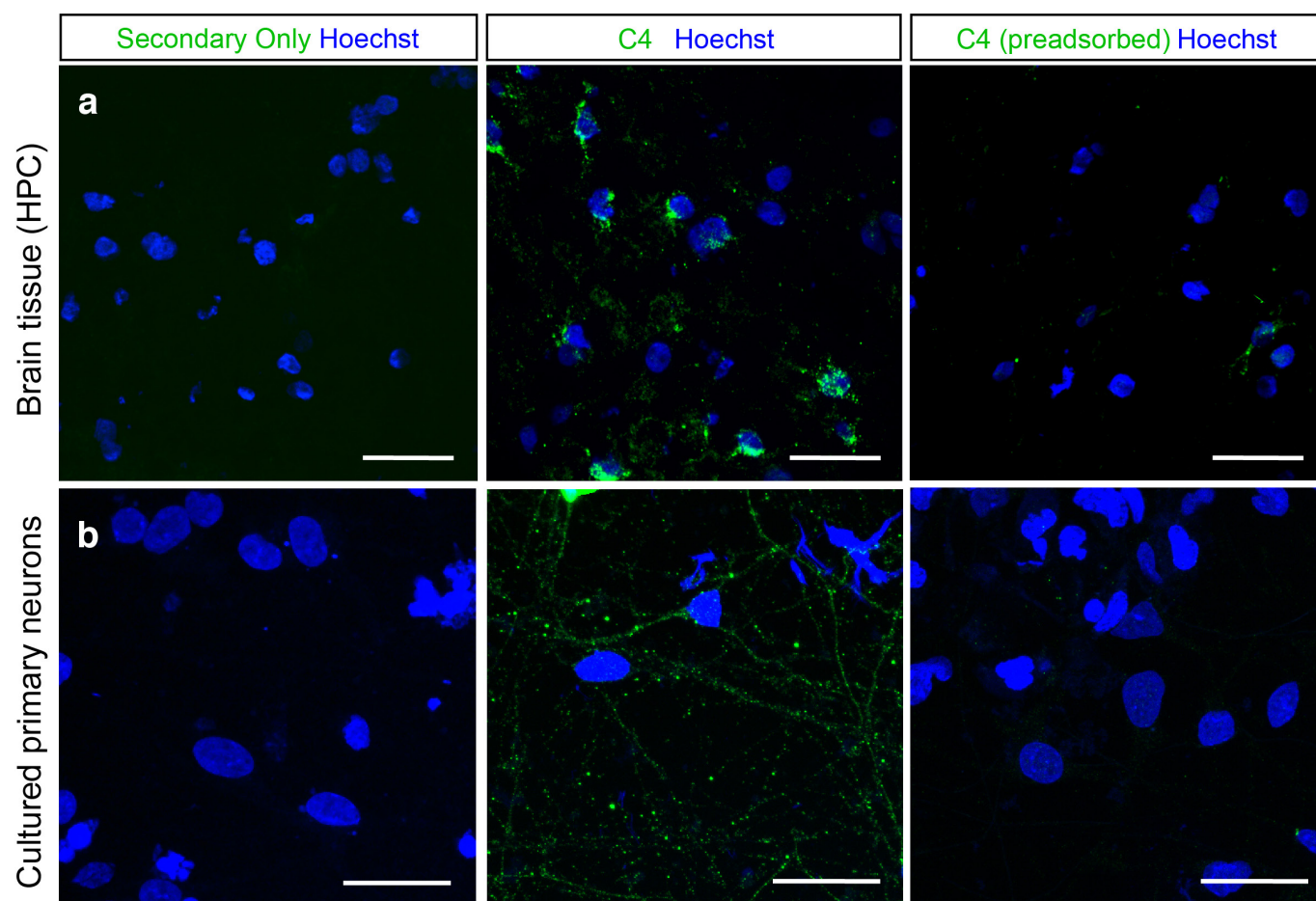
**Extended Data Figure 7 | Detailed association analysis for the most strongly associating classical *HLA* alleles.** The most strongly associating *HLA* loci were *HLA-B* (in primary analyses, Fig. 4a and Extended Data Fig. 6a) and *HLA-DRB1* and *HLA-DQB1* (in analyses controlling for the signal defined by rs13194504, Fig. 4c and Extended Data Fig. 6b). At these loci, the most strongly associating classical *HLA* alleles were *HLA-B\*0801*, *HLA-DRB1\*0301*, and *HLA-DQB\*02*, respectively. These *HLA* alleles are all in strong but partial LD with *C4* BS, the most protective of the *C4* alleles; they are also in partial LD with the low-risk allele at rs13194505,

representing the distinct signal several megabases to the left (Fig. 4). In joint analyses with each of these *HLA* alleles, genetically predicted *C4A* expression and rs13194505 continued to associate strongly with schizophrenia, while the *HLA* alleles did not. In further joint analyses with rs13194504 and genetically predicted *C4A* expression, 0 of 2,514 tested *HLA* SNP, amino acid and classical-allele polymorphisms (from ref. 55, including all variants with minor allele frequency (MAF) >0.005) associated with schizophrenia as strongly as rs13194504 or predicted *C4A* expression did.

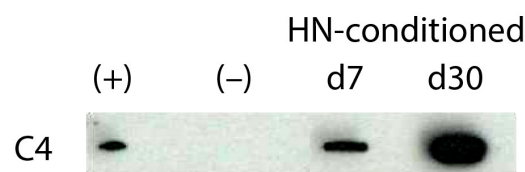


**Extended Data Figure 8 | Expression of *C4A* RNA in brain tissue (five brain regions) from 35 schizophrenia cases and 70 non-schizophrenia controls, from the Stanley Medical Research Institute Array Consortium. *C4A* RNA expression levels were measured by ddPCR. *P* values are derived from Mann–Whitney *U*-test.**



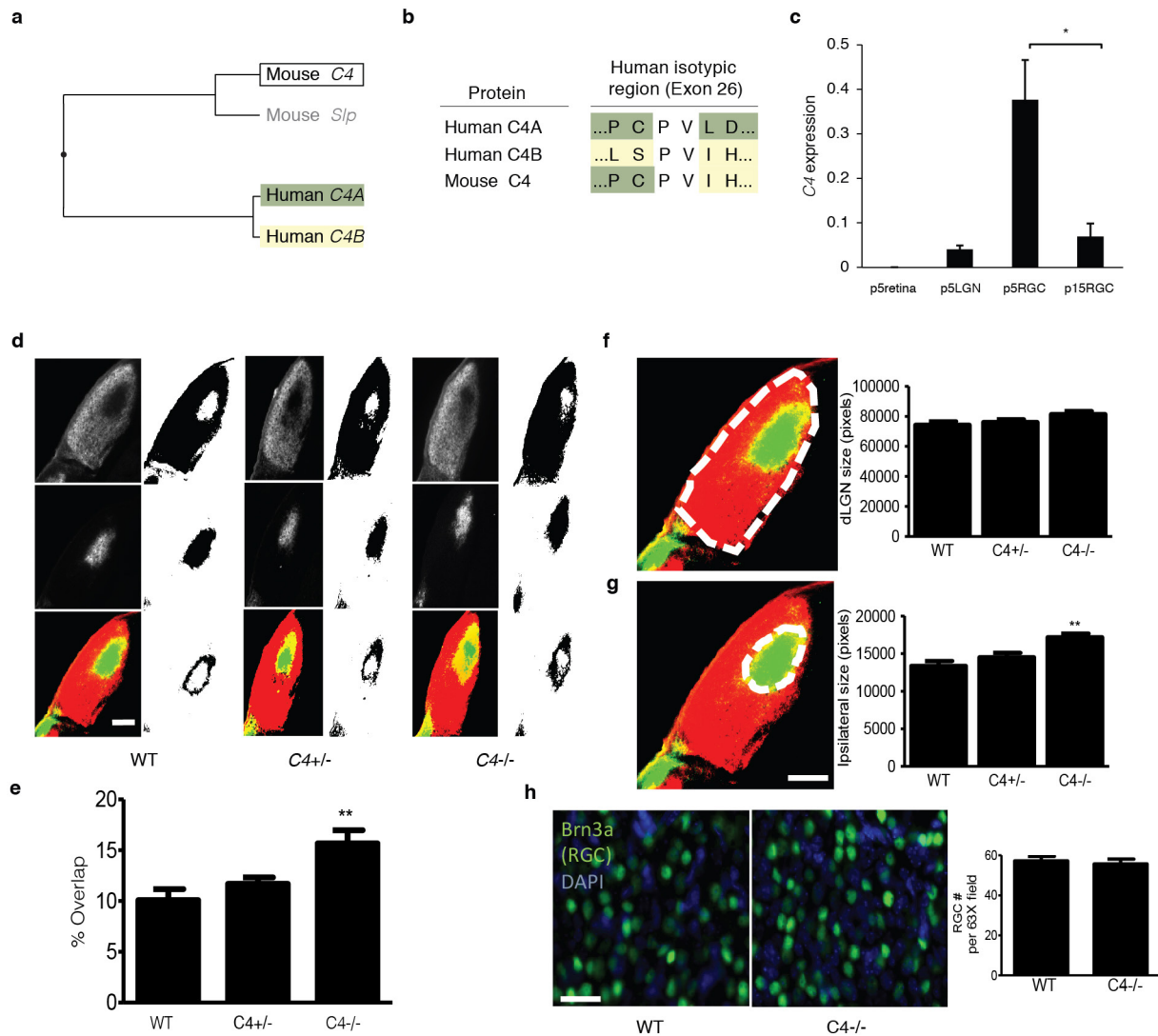


**c**



**Extended Data Figure 9 | Secretion of C4, and specificity of the monoclonal anti-C4 antibody for C4 protein in human brain tissue and cultured primary cortical neurons.** **a**, Brain tissue (from an individual affected with schizophrenia) was stained with a fluorescent secondary antibody, C4 antibody, or C4 antibody that was pre-adsorbed with purified C4 protein. Confocal images demonstrate the loss of immunoreactivity in the secondary-only and pre-adsorbed conditions. **b**, Primary human neurons were stained with a fluorescent secondary antibody, C4 antibody or C4 antibody that was pre-adsorbed with purified C4 protein. Confocal

images demonstrate the loss of immunoreactivity in the secondary-only and pre-adsorbed conditions. Scale bars, 25  $\mu$ m. **c**, Secretion of C4 protein by cultured primary neurons. Western blot for C4 protein analysis. (+) Purified human C4 protein. (-) Unconditioned medium, a negative control. HN-conditioned shows the same medium after conditioning by cultured human neurons at days 7 (d7) and 30 (d30). Details of western blot protocol, antibody catalogue numbers and concentrations used are in Supplementary Methods. C4 molecular weight, ~210 kDa.



**Extended Data Figure 10 | Mouse *C4* genes and additional analyses of the dLGN eye segregation phenotype in *C4* mutant mice and wild-type and heterozygous littermate controls.** **a**, The functional specialization of *C4* into *C4A* and *C4B* in humans does not have an analogy in mice. Although the mouse genome contains both a *C4* gene and a *C4*-like gene (classically called *Slp*), and these genes are also present as a tandem duplication within the mouse MHC locus, analysis of the encoded protein sequences indicates a distinct specialization, as illustrated by the protein phylogenetic tree. Top, mouse *Slp* is indicated in grey to reflect its potential pseudogenization: *Slp* is already known to have mutations at a C1s cleavage site, which are thought to abrogate activation of the protein through the classical complement pathway<sup>56</sup>; and the *M. musculus* reference genome sequence (mm10) at *Slp* shows a 1-bp deletion (relative to *C4*) within the coding region at chr17:34815158, which would be predicted to cause a premature termination of the encoded protein. In some genome data resources, mouse *Slp* and *C4* have been annotated respectively as '*C4a*' (for example, NM\_011413.2) and '*C4b*' (for example, NM\_009780.2) based on synteny with the human *C4A* and *C4B* genes, but the above sequence analysis indicates that they are not paralogous to *C4A* and *C4B*. **b**, Sequence differences between *C4A* and *C4B*—which are otherwise 99.5% identical at an amino acid level—are concentrated at the 'isotypic site' where they shape each isotype's relative affinity for different molecular targets<sup>19,20</sup>. At the isotypic site, mouse *C4* contains a combination of the residues present in human *C4A* and *C4B*. **c**, Expression

of mouse *C4* mRNA in whole retina and lateral geniculate nucleus (LGN) from P5 animals and in purified retinal ganglion cells (RGCs) from P5 and P15 animals. These time points were chosen as P5 is a time of more robust synaptic refinement in the retinogeniculate system compared to P15. The same assays detected no *C4* RNA in control RNA isolated from *C4*<sup>-/-</sup> mice (not shown). *n* = 3 samples for p5 retina, LGN, and P15 RGCs; *n* = 4 samples for P5 RGCs; \**P* < 0.05 by ANOVA with post-hoc Tukey–Kramer multiple-comparisons test. **d**, Representative images of dLGN innervation by contralateral projections (red in bottom image), ipsilateral projections (green in bottom image), and their overlap (yellow in bottom image). Scale bar, 100  $\mu$ m. **e**, Quantification of the percentage of total dLGN area receiving both contralateral and ipsilateral projections shows a significant increase in *C4*<sup>-/-</sup> compared to wild-type littermates (ANOVA, *n* = 5 mice per group, *P* < 0.01). These data are consistent with results using *R* value analysis as shown in Fig. 7. **f**, Quantification of total dLGN area showed no significant difference between wild-type and *C4*<sup>-/-</sup> mice (ANOVA, *n* = 5 per group, *P* > 0.05). **g**, Quantification of dLGN area receiving ipsilateral innervation showed a significant increase in ipsilateral territory in the *C4*<sup>-/-</sup> mice compared to wild-type littermates (ANOVA, *n* = 5 mice per group, *P* > 0.01). This result is consistent with defects in eye specific segregation. Scale bar, 100  $\mu$ m. **h**, The number of RGCs in the retina was estimated by counting the number of Brn3a<sup>+</sup> cells in wild-type and *C4*<sup>-/-</sup> mice. No differences were observed between wild-type and *C4*<sup>-/-</sup> mice (*t*-test, *n* = 4 mice per group, *P* > 0.05). Scale bar, 100  $\mu$ m.

# Naturally occurring p16<sup>Ink4a</sup>-positive cells shorten healthy lifespan

Darren J. Baker<sup>1</sup>, Bennett G. Childs<sup>2</sup>, Matej Durik<sup>1</sup>, Melinde E. Wijers<sup>1</sup>, Cynthia J. Sieben<sup>2</sup>, Jian Zhong<sup>1</sup>, Rachel A. Saltness<sup>1</sup>, Karthik B. Jeganathan<sup>1</sup>, Grace Casacang Verzosa<sup>3</sup>, Abdulmohammad Pezeshki<sup>4</sup>, Khashayarsha Khazaie<sup>4</sup>, Jordan D. Miller<sup>3</sup> & Jan M. van Deursen<sup>1,2</sup>

Cellular senescence, a stress-induced irreversible growth arrest often characterized by expression of p16<sup>Ink4a</sup> (encoded by the *Ink4a/Arf* locus, also known as *Cdkn2a*) and a distinctive secretory phenotype, prevents the proliferation of preneoplastic cells and has beneficial roles in tissue remodelling during embryogenesis and wound healing. Senescent cells accumulate in various tissues and organs over time, and have been speculated to have a role in ageing. To explore the physiological relevance and consequences of naturally occurring senescent cells, here we use a previously established transgene, *INK-ATTAC*, to induce apoptosis in p16<sup>Ink4a</sup>-expressing cells of wild-type mice by injection of AP20187 twice a week starting at one year of age. We show that compared to vehicle alone, AP20187 treatment extended median lifespan in both male and female mice of two distinct genetic backgrounds. The clearance of p16<sup>Ink4a</sup>-positive cells delayed tumorigenesis and attenuated age-related deterioration of several organs without apparent side effects, including kidney, heart and fat, where clearance preserved the functionality of glomeruli, cardio-protective K<sub>ATP</sub> channels and adipocytes, respectively. Thus, p16<sup>Ink4a</sup>-positive cells that accumulate during adulthood negatively influence lifespan and promote age-dependent changes in several organs, and their therapeutic removal may be an attractive approach to extend healthy lifespan.

Cellular senescence is a well-established cancer defence mechanism that has also been proposed to have roles in ageing and age-associated diseases, presumably through the depletion of stem and progenitor cells, and the adverse actions of the senescence-associated secretory phenotype, which consists of many proinflammatory cytokines and chemokines, matrix metalloproteinases and growth factors<sup>1–3</sup>. Consistent with this idea is the observation that interference with senescent cell accumulation in BubR1 progeroid mice delays several of the ageing-associated disorders that these animals develop<sup>4,5</sup>. However, because progeroid syndromes do not mimic the complex degenerative changes of ageing completely, the relevance of these findings has remained unclear. Furthermore, recent studies showing that senescent cells have beneficial effects in injury repair and tissue remodelling<sup>6–10</sup> have called into question the simplistic view of senescence as only a driver of age-dependent pathologies, raising the specter that senescent cell clearance might remove useful cells in addition to detrimental ones. Here we investigated the identity and physiological effect of naturally occurring senescent cells using *INK-ATTAC* (hereafter termed *ATTAC*)<sup>4</sup>, a transgenic mouse model that expresses the FK506-binding-protein–caspase 8 (FKBP–Casp8) fusion protein and green fluorescent protein (GFP) under the control of a minimal *Ink4a* (also known as *Ink4a/Arf* or *Cdkn2a*) promoter fragment transcriptionally active in senescent cells<sup>4,11</sup>. Earlier we have shown that, in BubR1 progeroid mice, *ATTAC* ablates p16<sup>Ink4a</sup>-positive senescent cells upon administration of AP20187 (AP), a dimerizer that activates FKBP-fused Casp8 (ref. 4). Our first objective was to validate the properties of *ATTAC* in naturally occurring p16<sup>Ink4a</sup>-positive senescent cells.

## *ATTAC* clears senescent adipocyte progenitor cells

Our initial validation was focused on fat. We collected GFP<sup>+</sup> and GFP<sup>−</sup> cell populations from inguinal white adipose tissue (iWAT) of

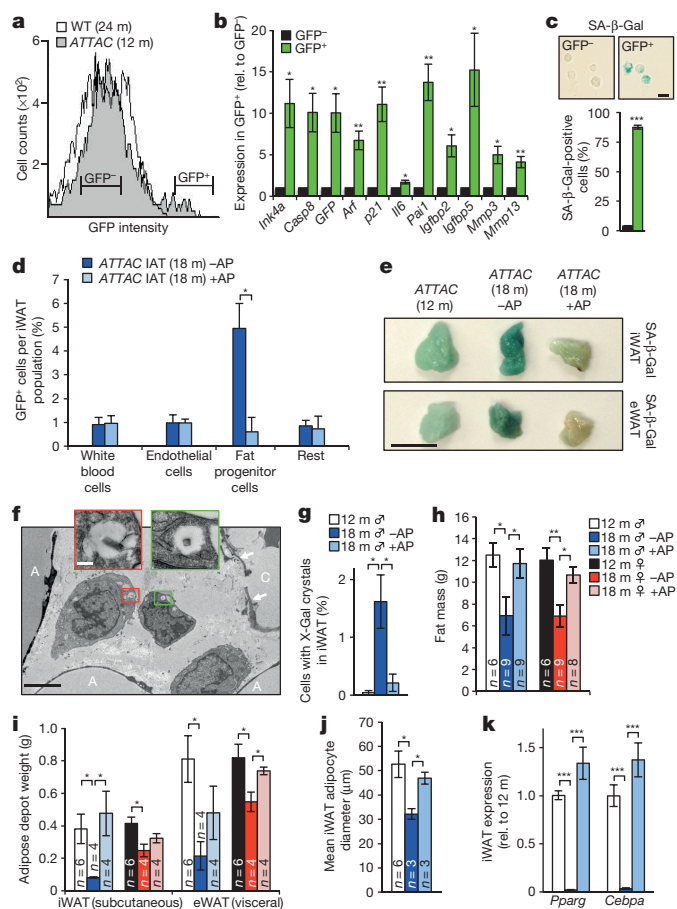
12-month-old *ATTAC* mice by FACS (Fig. 1a). GFP<sup>+</sup> cells expressed much higher levels of *Ink4a* and *FKBP-Casp8* than GFP<sup>−</sup> cells, as well as a broad panel of senescence markers (Fig. 1b). GFP<sup>+</sup> cells, but not GFP<sup>−</sup> cells, were also highly positive for senescence-associated-β-galactosidase (SA-β-Gal; Fig. 1c). Furthermore, intact iWAT from aged, but not young *ATTAC* mice had SA-β-Gal activity, but less than iWAT of BubR1 progeroid mice, a difference also reflected in *Ink4a*, *FKBP-Casp8* and GFP transcript levels (Extended Data Fig. 1a, b).

iWAT of 18-month-old *ATTAC* mice treated bi-weekly with AP from 12 months onwards had eightfold less GFP<sup>+</sup> adipocyte progenitors than vehicle-injected controls, although total progenitor cell numbers remained unchanged (Fig. 1d and Extended Data Fig. 1c). SA-β-Gal staining and quantitative reverse transcriptase PCR (qRT-PCR) analysis of senescence markers confirmed that *Ink4a*-positive senescent cells in iWAT increased between 12 and 18 months, and that AP eliminated most of these cells (Fig. 1e and Extended Data Fig. 1d). Consistent with senescence of progenitor cells, transmission electron microscopy (TEM) on SA-β-Gal-stained iWAT showed that X-Gal crystals were present in small perivascular cells rather than endothelium, white blood cells or adipocytes (Fig. 1f). X-Gal crystals were found in 0.2% and 1.6% of total iWAT cells from AP-treated and control mice, respectively (Fig. 1g).

Clearance of *Ink4a*-positive cells prevented loss of fat mass occurring between 12 and 18 months (Fig. 1h, i and Extended Data Fig. 1e). Age-dependent fat tissue dysfunction is characterized by decreased adipogenesis and adipocyte atrophy<sup>12</sup>. Consistent with this, adipocyte size decreased between 12 and 18 months of age (Fig. 1j), as did transcript levels of two key transcriptional regulators of adipogenesis, *Pparg* and *Cebpa* (Fig. 1k). AP treatment of *ATTAC* mice prevented these decreases. Collectively, these data indicate that senescence contributes to age-dependent fat tissue alterations.

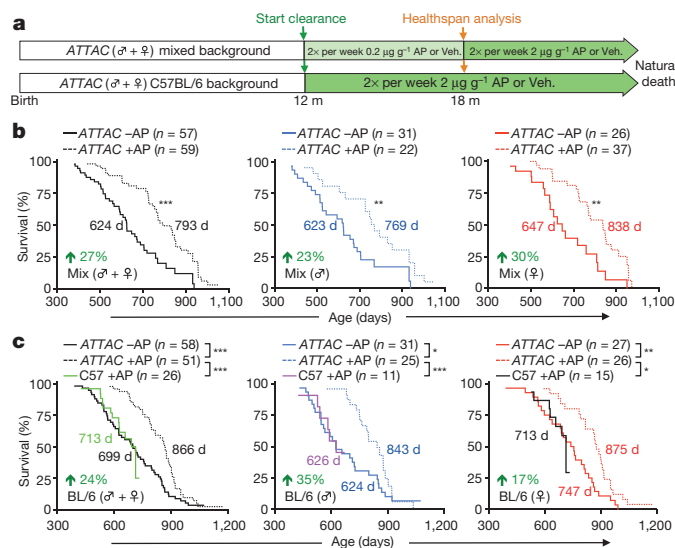
<sup>1</sup>Department of Pediatric and Adolescent Medicine, Mayo Clinic College of Medicine, Rochester, Minnesota 55905, USA. <sup>2</sup>Department of Biochemistry and Molecular Biology, Mayo Clinic College of Medicine, Rochester, Minnesota 55905, USA. <sup>3</sup>Division of Cardiovascular Surgery, Mayo Clinic College of Medicine, Rochester, Minnesota 55905, USA. <sup>4</sup>Department of Immunology, Mayo Clinic College of Medicine, Rochester, Minnesota 55905, USA.





**Figure 1 | Clearance of senescent fat progenitor cells attenuates age-related lipodystrophy.** **a**, FACS profiles of single-cell suspensions from iWAT of the indicated 12- and 24-month-old (m) mice. **b**, GFP<sup>+</sup> and GFP<sup>-</sup> cell populations from iWAT of 12-month-old ATTAC mice (see **a** for sorting brackets) analysed by qRT-PCR ( $n = 6$  mice). *p21* is also known as *Cdkn1a*; *Pa1* is also known as *Serpine1*. **c**, SA-β-Gal activity in GFP<sup>+</sup> and GFP<sup>-</sup> iWAT cells ( $n = 3$  mice). **d**, GFP<sup>+</sup> cells in the indicated iWAT (IAT) cell populations ('rest' represents the iWAT vascular stromal fraction minus leukocytes, endothelial cells and progenitors). **e–i**, Fat-related analyses of C57BL/6 ATTAC mice before treatment (12 months) or after 6 months of treatment with vehicle (18 months –AP) or AP (18 months +AP). **e**, SA-β-Gal activity in iWAT and epididymal WAT (eWAT). **f**, Electron micrograph showing perivascular X-Gal-positive cells from an 18-month-old vehicle-treated C57BL/6 ATTAC male. A, adipocyte; C, capillary. Arrows mark endothelial cells. **g**, Quantification of iWAT cells containing X-Gal crystals ( $n = 4$  mice per treatment). **h**, Fat mass measurements. **i**, iWAT and eWAT depot weights. **j**, Mean adipocyte diameters in iWAT. **k**, Expression of adipogenesis markers in iWAT ( $n = 4$  mice per group). Scale bars, 10 μm (**c**), 0.5 cm (**e**), 2 μm (**f**) and 200 nm (**f**, inset). Error bars indicate s.e.m. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  (one-sample *t*-tests using a theoretical mean of 1 (**b**), and unpaired two-tailed *t*-tests (**c**, **d**, **g–k**)).

Adipose tissue of young ATTAC mice lacked SA-β-Gal activity but contained p16<sup>Ink4a</sup> (Extended Data Fig. 1a, f–h). This p16<sup>Ink4a</sup> pool did not decline after AP treatment. Similar results were obtained with early passage ATTAC MEFs (Extended Data Fig. 1i–k), indicating that baseline *Ink4a* promoter activity in non-senescent cells is insufficient for FKBP–Casp8-mediated apoptosis. Furthermore, ATTAC lacks *Ink4a* promoter elements required for transgene induction in replicating pre-neoplastic cells that robustly express endogenous *Ink4a* owing to Rb inactivation and AP fails to kill these cells (Extended Data Fig. 2a–g). ATTAC was also not induced in peripheral blood T lymphocytes that robustly engage endogenous *Ink4a* with ageing without concomitant expression of several senescence markers<sup>2,13</sup> (Extended Data Fig. 2h),



**Figure 2 | Senescent cell clearance extends lifespan.** **a**, Study design for clearance of senescent cells in mixed and C57BL/6 mouse cohorts. Healthspan analysis was done at 18 months, an age at which relatively few mice in vehicle (Veh.)- or AP-treated groups have died, and bias owing to selection for long-lived animals is unlikely. **b**, **c**, Survival curves for vehicle-treated (–AP) and AP-treated (+AP) mixed (**b**) and C57BL/6 (**c**) mice. Median survival (in days, d) and percentage increase in median survival are indicated. We note that median lifespans of our vehicle-treated cohorts are similar to those of wild-type mice administered AP (**c**). \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  (log-rank tests).

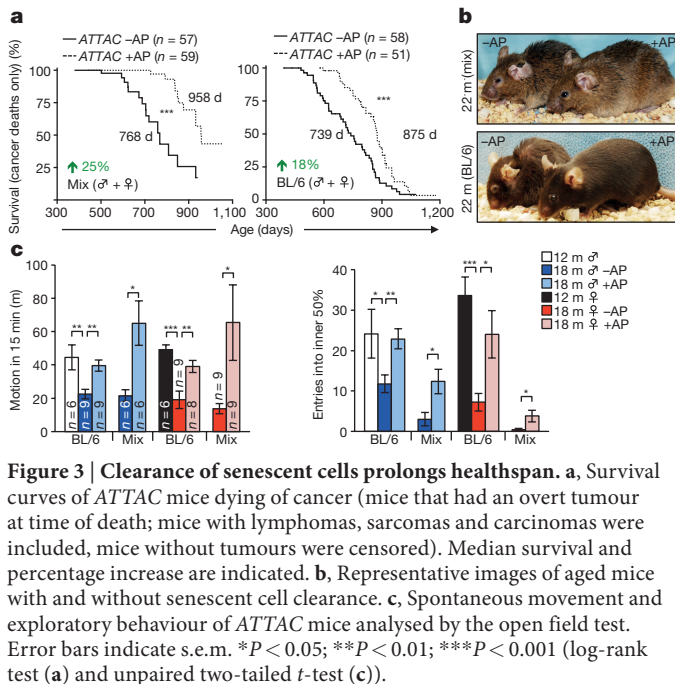
further indicating that transgene induction is selective for senescence. However, these limited analyses certainly do not exclude the possibility that other *Ink4a*-positive non-senescent cells engage ATTAC and die after AP exposure (Supplementary Information).

### Clearance by ATTAC is partial and tissue selective

To extend our analysis of the properties of ATTAC, we analysed the transcript levels of *Ink4a*, *FKBP–Casp8*, GFP and a panel of senescence markers in a broad spectrum of tissues of 2-, 12- and 18-month-old untreated ATTAC mice, including skeletal muscle, eye, kidney, lung, heart, liver, colon and spleen. ATTAC mice treated with AP between 12 and 18 months were included to assess senescent cell clearance rates. *Ink4a* expression markedly increased between 12 and 18 months in all tissues examined, which coincided with the induction of *FKBP–Casp8*, GFP and several senescence markers (Extended Data Fig. 3a). Increased expression of these transcripts was blunted to varying degrees by AP treatment in all tissues examined except colon and liver, indicating that the ATTAC system eliminates *Ink4a*-positive senescent cells in a partial and tissue/organ selective fashion. Blunting of age-dependent increases in *Il6*, *Il1a* and *Tnfa* expression in fat, skeletal muscle and kidney by AP treatment (Extended Data Fig. 3b) suggests that *Ink4a*-positive cell clearance ameliorates inflammation in these tissues.

### p16<sup>Ink4a</sup> cells shorten lifespan and healthspan

To examine the effect of p16<sup>Ink4a</sup>-positive cell clearance on health and lifespan, we sequentially established two cohorts of ATTAC transgenic mice (Fig. 2a). The initial cohort was on a C57BL/6-129Sv-FVB mixed genetic background fed a diet containing 9% fat. We note that this diet shortens lifespan compared to diets with 5% fat typically used in lifespan studies (Extended Data Fig. 4a, b and Supplementary Information). The later cohort was on a congenic C57BL/6 background fed a standard 5% fat diet. At 12 months of age, when p16<sup>Ink4a</sup> cells started to accumulate in several tissues (Extended Data Figs 1d and 3a), mice were injected twice a week with AP or vehicle until they became moribund or died of natural causes. Mice separate from the longevity cohorts were examined for a series of age-sensitive outcomes at



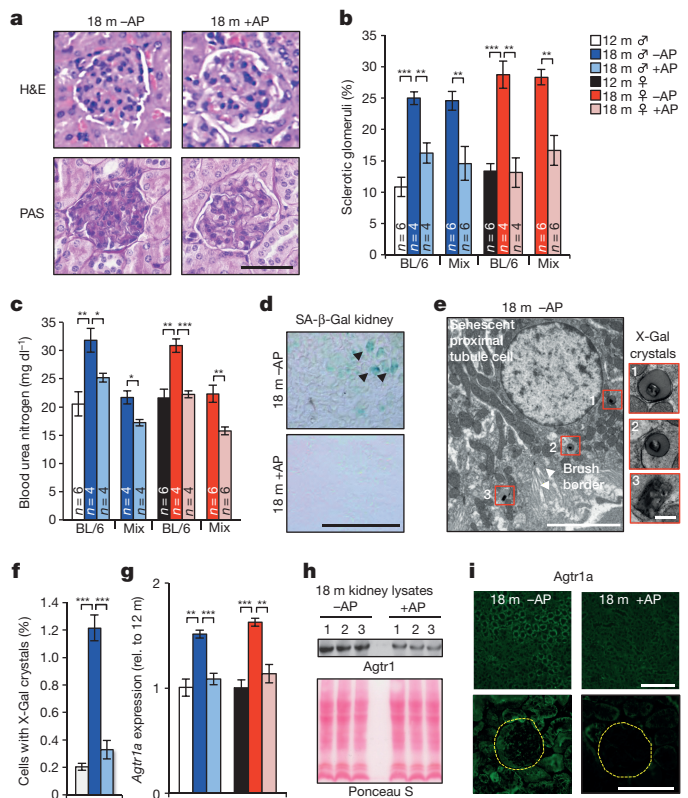
**Figure 3 | Clearance of senescent cells prolongs healthspan.** **a**, Survival curves of *ATTAC* mice dying of cancer (mice that had an overt tumour at time of death; mice with lymphomas, sarcomas and carcinomas were included, mice without tumours were censored). Median survival and percentage increase are indicated. **b**, Representative images of aged mice with and without senescent cell clearance. **c**, Spontaneous movement and exploratory behaviour of *ATTAC* mice analysed by the open field test. Error bars indicate s.e.m. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  (log-rank test (**a**) and unpaired two-tailed *t*-test (**c**)).

18 months, an age at which relatively few mice in each of the cohorts had died. Data for both sexes combined showed that median lifespans of mixed and C57BL/6 AP-treated animals were increased by 27% and 24%, respectively. Median lifespans for each sex separately were also significantly extended in AP-treated cohorts irrespective of genetic background, with increases ranging from 17% to 35% (Fig. 2b, c).

Maximum lifespan<sup>14</sup> was significantly increased for mixed AP-treated males and females combined ( $P = 0.0295$ ), but not for females and males individually. Maximum lifespan was not extended for C57BL/6 AP-treated animals, either combined or separately. Importantly, AP treatment of mice lacking the *ATTAC* transgene did not improve lifespan (Fig. 2c). We note that the median lifespan of vehicle-treated C57BL/6 males, but not females, was short of the normal range of lifespans for unmanipulated males of this strain at different laboratories<sup>15–28</sup> (Extended Data Fig. 4c, d), suggesting that repetitive vehicle injection stress may have negatively affected C57BL/6 male longevity (Supplementary Information).

In both cohorts, AP treatment had no effect on the incidence or spectrum of macroscopically detectable tumours at autopsy, although tumour latency was increased (Fig. 3a and Extended Data Fig. 5a–d). Median lifespan extensions of AP-treated mice dying without tumours ranged from 24% to 42% (Extended Data Fig. 5e), indicating that increased longevity was not merely due to a tumour-protective effect. AP-treated mice were overtly indistinguishable from vehicle-injected littermates at 18 months of age, but typically had a healthier appearance by 22 months (Fig. 3b). AP treatment delayed cataract formation in both males and females on a C57BL/6 background (Extended Data Fig. 5f, g). Despite a lack of overt difference at 18 months, AP-treatment prevented age-dependent reductions in both spontaneous activity and exploratory behaviour measured by open-field testing (Fig. 3c), which was independent of sex and genetic background.

Extended tests on these mice showed no differences in motor coordination and balance, memory, exercise ability or muscle strength (Extended Data Fig. 6), indicating that 6 months of p16<sup>Ink4a</sup>-positive cell clearance did not affect these age-sensitive traits. Similarly, age-related changes in the circulating haematopoietic profile, including regulatory T-cell levels, were also not affected (Extended Data Fig. 7). At 18 months, fasting glucose levels, glucose tolerance, insulin sensitivity and circulating Igf1 levels were not significantly different in AP- and vehicle-treated C57BL/6 *ATTAC* females, as was signalling downstream



**Figure 4 | Senescent cells cause glomerulosclerosis, kidney dysfunction and renal RAAS hyperactivity.** **a**, Images of sclerotic (left) and normal (right) glomeruli from the indicated mice. H&E, haematoxylin and eosin; PAS, periodic acid–Schiff. **b**, Quantification of sclerotic glomeruli. **c**, Measurements of blood urea nitrogen levels. **d**, SA-β-Gal-stained kidney sections. **e**, Electron micrograph showing a X-Gal crystal-containing renal epithelial cell with brush border membrane (arrowheads). Insets show X-Gal crystal close-ups. **f**, Percentage of cells with X-Gal crystals in renal sections ( $n = 5$  TEM grids for each treatment group). **g**, Renal expression of *Agtr1a* analysed by qRT-PCR ( $n = 4$  mice per group). **h**, Western blot of kidney lysates probed for *Agtr1a* ( $n = 3$  mice per treatment group). Ponceau S staining served as loading control. **i**, Immunostaining of kidney sections for *Agtr1a*. Yellow circles denote glomeruli. Scale bars, 50  $\mu$ m (**a** and **i**, bottom), 250  $\mu$ m (**d**), 5  $\mu$ m (**e**), 200 nm (**e**, insets) and 100  $\mu$ m (**i**, top). Error bars indicate s.e.m. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  (unpaired two-tailed *t*-tests). For gel source data, see Supplementary Fig. 1. All mice in **d–i** were C57BL/6 *ATTAC*.

of Igf1 and insulin receptors in fat, kidney and muscle, three tissues where we observed clearance of p16<sup>Ink4a</sup> cells (Extended Data Fig. 8).

### Renal epithelial cell senescence and glomerulosclerosis

To further investigate the effect of p16<sup>Ink4a</sup> cells on physiological functions that change with age<sup>29</sup>, we focused on kidney and heart, two vital organs in which we observed *ATTAC*-mediated clearance. Aged kidneys are characterized by the formation of sclerotic glomeruli, which affect glomerular filtration rates, impair kidney function, and lead to increased blood urea nitrogen levels<sup>30,31</sup>. Indeed, glomerulosclerosis significantly increased between 12 and 18 months in vehicle-treated *ATTAC* mice. AP treatment markedly reduced glomerulosclerosis independent of sex and genetic background (Fig. 4a, b), which correlated with attenuated age-related increases in blood urea nitrogen (Fig. 4c), indicating preserved kidney function. SA-β-Gal staining of kidney sections confirmed that AP-mediated disposal of senescent cells was quite efficient (Fig. 4d). By TEM, we observed X-Gal crystals in 0.3% and 1.2% of renal cells of 18-month-old treated and untreated mice, respectively (Fig. 4e, f). Surprisingly, SA-β-Gal<sup>+</sup> cells were not located in the glomeruli but in proximal tubules (Fig. 4e), raising the

question as to how senescent tubular brush-border epithelial cells might promote glomerulosclerosis.

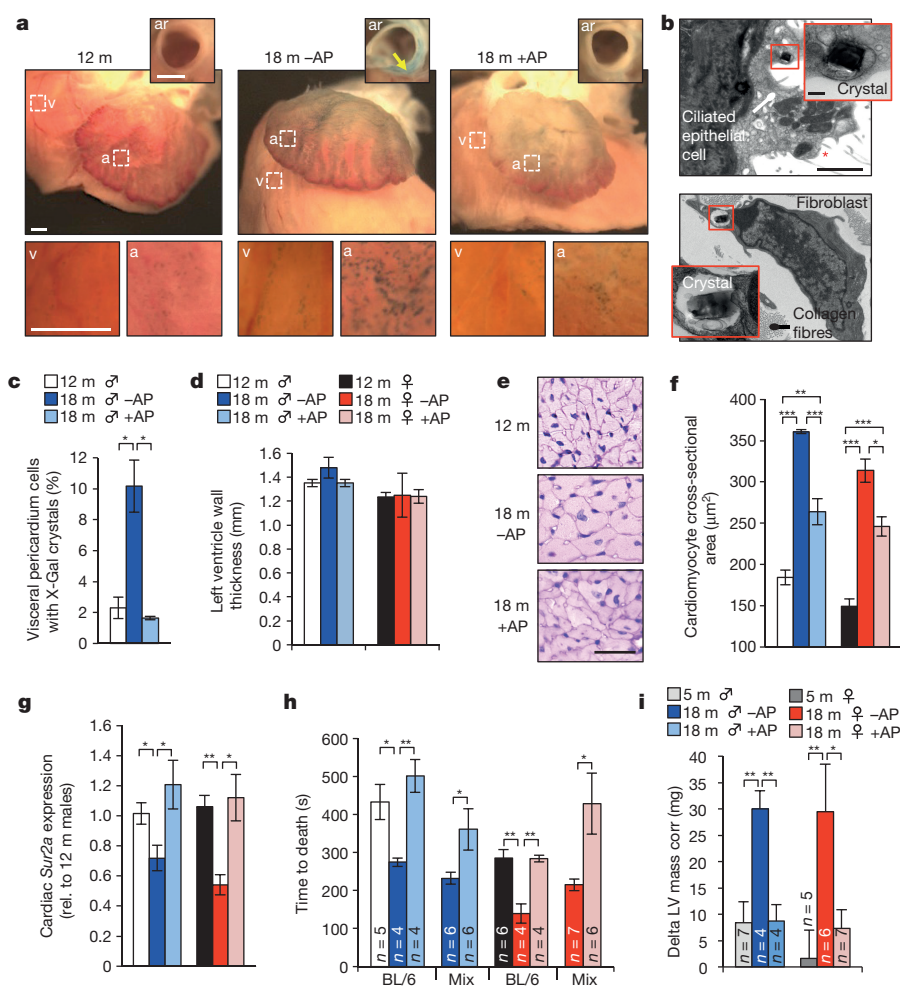
Angiotensin receptor blockers attenuate age-related glomerulosclerosis under normotensive conditions, which has led to the idea that overactivation of the local renin-angiotensin-aldosterone system (RAAS) drives the formation of sclerotic glomeruli<sup>31,32</sup>. Consistent with this, kidney transcript levels of a key component of this system, angiotensin receptor 1a (*Agtr1a*), increased between 12 and 18 months (Fig. 4g). By contrast, no such increase was observed in AP-treated mice. Western blotting confirmed that *Agtr1a* protein levels were lower in AP-treated kidney samples (Fig. 4h). Both renal tubules and glomeruli contributed to this decline, as demonstrated by immunolabelling of kidney sections for *Agtr1a* (Fig. 4i). These data suggest that senescent renal epithelial cells may produce senescence-associated secretory phenotype components that hyperactivate the local RAAS in kidney.

### p16<sup>Ink4a</sup> cells contribute to cardiac ageing

Hearts of 12-month-old mice showed SA- $\beta$ -Gal activity at the atrial and ventricular surface (Fig. 5a), which increased with ageing in vehicle-treated but not AP-treated mice. TEM revealed that ciliated epithelial cells and fibroblasts in the pericardium contained X-Gal crystals

(Fig. 5b, c). We also observed SA- $\beta$ -Gal<sup>+</sup> smooth muscle cells in the aortic root wall of vehicle-treated animals (Fig. 5a and Extended Data Fig. 9a). At 18 months, vehicle- and AP-treated *ATTAC* mice analysed by echocardiography showed no significant differences in heart rate, left ventricular mass, thickness and diameter, ejection fraction and fractional shortening, all of which were also unchanged from 12-month-old mice (Extended Data Fig. 9b–g). Cardiac ageing at the histological level is characterized by a loss of ventricular cardiomyocytes due to decreasing ability to replace apoptotic or necrotic cardiomyocytes, inducing hypertrophy of the remaining cardiomyocytes<sup>33</sup>. Morphometric analysis of heart sections showed that ventricular wall thickness was the same in vehicle- and AP-treated animals, regardless of sex (Fig. 5d). Ventricular cardiomyocytes of AP-treated mice, however, were much smaller (Fig. 5e, f), suggesting that they had more cardiomyocytes than vehicle-treated mice. Taken together, these data suggest that p16<sup>Ink4a</sup>-positive cells are key drivers of this age-related cardiac phenotype.

Cardiac ageing is also characterized by decreasing stress tolerance, which has been attributed to decreasing numbers of ATP-sensitive potassium ( $K_{ATP}$ ) channels in the sarcolemma of cardiac myocytes due to age-related reductions in the expression of *Sur2a*, a key



**Figure 5 | Senescent cells promote age-related cardiomyocyte hypertrophy and loss of cardiac stress tolerance.** **a**, SA- $\beta$ -Gal-stained hearts. Insets show aortic roots (ar) from a transverse plane (arrow marks the aortic root wall) or close-ups of the ventricular (v) and arterial (a) boxed areas. **b**, Electron micrographs of X-Gal-positive cells in the pericardium (red asterisk marks cilia). Insets show close-ups of X-Gal crystals. **c**, Quantification of cells with X-Gal crystals in the visceral pericardium ( $n = 4$  mice per treatment). **d**, Measurements of left ventricle wall thickness ( $n = 4$  mice per group). **e**, Representative cardiomyocyte

cross-sectional images ( $n = 4$  mice per group). **f**, Quantification of **e**. **g**, Analysis of *Sur2a* expression in hearts by qRT-PCR ( $n = 4$  mice per group). **h**, Cardiac stress resistance determined by measuring the time to death after injection of a lethal dose of isoproterenol. **i**, Change in left ventricular (LV) mass in response to sublethal doses of isoproterenol ( $10 \text{ mg kg}^{-1}$ ) after ten doses administered over 5 days. Scale bars, 1 mm (**a**),  $2 \mu\text{m}$  (**b**) and  $200 \text{ nm}$  (**b**, inset). Error bars indicate s.e.m. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  (unpaired two-tailed *t*-tests). All mice, except for **h**, were C57BL/6 *ATTAC*.



regulatory subunit of K<sub>ATP</sub> channels<sup>34</sup>. Cardiac *Sur2a* (also known as *Abcc9*) expression indeed decreased between 12 and 18 months in vehicle-treated mice (Fig. 5g), but no such decline was observed in AP-treated mice, suggesting that cardiac stress tolerance was preserved. To test this, we measured the time to death from arrhythmogenesis after injection of 580 mg kg<sup>-1</sup> of the  $\beta$ -adrenergic agonist isoproterenol<sup>35</sup>. Consistent with age-related loss of cardiac stress resistance, 18-month-old vehicle-treated mice died faster than 12-month-old mice (Fig. 5h). By contrast, the time to death was not accelerated in AP-treated mice. In a second, more physiological stress test, we conducted echocardiographic measurements of ventricular mass before and after eliciting cardiac stress over a 6-day period, by administering 10 mg kg<sup>-1</sup> isoproterenol twice a day. Cardiac mass significantly increased in 18-month-old vehicle-treated mice, whereas AP-treated mice were capable of handling the applied stress without such adaptive response, similar to young mice (Fig. 5i).

We also examined whether AP-treatment caused any off-target or detrimental effects. Wild-type mice lacking *ATTAC* did not show healthspan improvements when treated with AP (Extended Data Fig. 9h–l). Consistent with a beneficial role of senescent cells only for the duration of wound repair<sup>6–8</sup>, 18-month-old AP-treated *ATTAC* mice repaired cutaneous wounds when drug administration was suspended during healing (Extended Data Fig. 10a–c). When AP was administered during wound closure, healing was delayed with similar kinetics as 4- or 18-month-old *ATTAC* mice without previous AP treatment, indicating that acute senescence mechanisms are preserved with ageing and not influenced by constitutive clearance of senescent cells<sup>3</sup>. Furthermore, 18-month-old AP-treated mice showed no evidence of increased fibrosis in skin and other tissues, despite a role for senescent cells in limiting fibrosis during tissue repair (Extended Data Fig. 10d).

## Discussion

Here we eliminated p16<sup>Ink4a</sup> cells from non-progeroid mice using *ATTAC* to begin to address how senescent cells influence health and lifespan. We observe median lifespan extensions in AP-treated *ATTAC* mice on two distinct genetic backgrounds and diets, indicating that age-related accumulation of p16<sup>Ink4a</sup> cells negatively affects longevity. It will be useful to optimize senescent cell removal protocols and methods further because the longevity of male C57BL/6 mice seemed negatively affected by repetitive vehicle injection stress, and because clearance was partial and several key tissues were refractory to clearance, including liver and colon. This would be particularly relevant for further studies of maximum lifespan, extension of which is thought to be reflective of a more generalizable 'anti-ageing' effect<sup>29</sup> (Supplementary Information).

These methodological caveats aside, we find that removal of p16<sup>Ink4a</sup> cells from midlife on delays the progression of neoplastic disease irrespective of genetic background and diet. Clearance had no effect on the spectrum or incidence of post-mortem tumours, which is consistent with the hypothesis that senescent cells help create a tissue microenvironment that promotes tumour progression<sup>1</sup>. In addition, several age-dependent changes<sup>29</sup>, including alterations in fat, eye, heart and kidney, as well as the age-dependent decline in spontaneous activity and exploratory behaviour, occur more slowly during a six-month period of p16<sup>Ink4a</sup> cell clearance starting at midlife. Complementary mechanistic studies suggest that local p16<sup>Ink4a</sup> cells in fat, kidney and heart exert their effects through distinct mechanisms, involving progenitor cell dysfunction, RAAS overactivation, and *Sur2a* downregulation, respectively, although systemic effects cannot be excluded given the global nature of the clearance method applied here.

Thus, our data best fit a model in which p16<sup>Ink4a</sup> cells act to shorten healthy lifespan by promoting tumour progression and age-dependent changes that functionally impair certain tissues and organs, including vital organs such as kidney and heart. This, together with the key observation that elimination of p16<sup>Ink4a</sup> cells is not associated with any overt detrimental effects raises the possibility that this approach may

be useful to treat aspects of age-related functional decline, age-related diseases that involve senescent cells, or side effects of therapies that create senescent cells<sup>36</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 31 March; accepted 21 December 2015.**

**Published online 3 February 2016.**

- Campisi, J. Aging, cellular senescence, and cancer. *Annu. Rev. Physiol.* **75**, 685–705 (2013).
- Sharpless, N. E. & Sherr, C. J. Forging a signature of *in vivo* senescence. *Nature Rev. Cancer* **15**, 397–408 (2015).
- van Deursen, J. M. The role of senescent cells in ageing. *Nature* **509**, 439–446 (2014).
- Baker, D. J. *et al.* Clearance of p16<sup>Ink4a</sup>-positive senescent cells delays ageing-associated disorders. *Nature* **479**, 232–236 (2011).
- Baker, D. J. *et al.* Opposing roles for p16<sup>Ink4a</sup> and p19<sup>Arf</sup> in senescence and ageing caused by BubR1 insufficiency. *Nature Cell Biol.* **10**, 825–836 (2008).
- Jun, J. I. & Lau, L. F. The matricellular protein CCN1 induces fibroblast senescence and restricts fibrosis in cutaneous wound healing. *Nature Cell Biol.* **12**, 676–685 (2010).
- Krizhanovsky, V. *et al.* Senescence of activated stellate cells limits liver fibrosis. *Cell* **134**, 657–667 (2008).
- Demaria, M. *et al.* An essential role for senescent cells in optimal wound healing through secretion of PDGF-AA. *Dev. Cell* **31**, 722–733 (2014).
- Muñoz-Espín, D. *et al.* Programmed cell senescence during mammalian embryonic development. *Cell* **155**, 1104–1118 (2013).
- Storer, M. *et al.* Senescence is a developmental mechanism that contributes to embryonic growth and patterning. *Cell* **155**, 1119–1130 (2013).
- Wang, W., Wu, J., Zhang, Z. & Tong, T. Characterization of regulatory elements on the promoter region of p16<sup>Ink4a</sup> that contribute to overexpression of p16 in senescent fibroblasts. *J. Biol. Chem.* **276**, 48655–48661 (2001).
- Tchkonia, T. *et al.* Fat tissue, aging, and cellular senescence. *Ageing Cell* **9**, 667–684 (2010).
- Liu, Y. *et al.* Expression of p16<sup>Ink4a</sup> in peripheral blood T-cells is a biomarker of human aging. *Ageing Cell* **8**, 439–448 (2009).
- Wang, C., Li, Q., Redden, D. T., Weindruch, R. & Allison, D. B. Statistical methods for testing effects on "maximum lifespan". *Mech. Ageing Dev.* **125**, 629–632 (2004).
- Goodrick, C. L., Ingram, D. K., Reynolds, M. A., Freeman, J. R. & Cider, N. Effects of intermittent feeding upon body weight and lifespan in inbred mice: interaction of genotype and age. *Mech. Ageing Dev.* **55**, 69–87 (1990).
- Blackwell, B. N., Bucci, T. J., Hart, R. W. & Turturro, A. Longevity, body weight, and neoplasia in *ad libitum*-fed and diet-restricted C57BL/6 mice fed NIH-31 open formula diet. *Toxicol. Pathol.* **23**, 570–582 (1995).
- Bronson, R. T. & Lipman, R. D. Reduction in rate of occurrence of age related lesions in dietary restricted laboratory mice. *Growth Dev. Ageing* **55**, 169–184 (1991).
- Conti, B. *et al.* Transgenic mice with a reduced core body temperature have an increased life span. *Science* **314**, 825–828 (2006).
- Coschigano, K. T. *et al.* Deletion, but not antagonism, of the mouse growth hormone receptor results in severely decreased body weights, insulin, and insulin-like growth factor I levels and increased life span. *Endocrinology* **144**, 3799–3810 (2003).
- Forster, M. J., Morris, P. & Sohal, R. S. Genotype and age influence the effect of caloric intake on mortality in mice. *FASEB J.* **17**, 690–692 (2003).
- Ikeno, Y. *et al.* Housing density does not influence the longevity effect of calorie restriction. *J. Gerontol. A Biol. Sci. Med. Sci.* **60**, 1510–1517 (2005).
- Kunster, I. & Leuenberger, H. G. Gerontological data of C57BL/6J mice. I. Sex differences in survival curves. *J. Gerontol.* **30**, 157–162 (1975).
- Selman, C. *et al.* Evidence for lifespan extension and delayed age-related biomarkers in insulin receptor substrate 1 null mice. *FASEB J.* **22**, 807–818 (2008).
- Sohal, R. S., Ku, H. H., Agarwal, S., Forster, M. J. & Lal, H. Oxidative damage, mitochondrial oxidant generation and antioxidant defenses during aging and in response to food restriction in the mouse. *Mech. Ageing Dev.* **74**, 121–133 (1994).
- Srivastava, V. K., Tilley, R. D., Hart, R. W. & Busbee, D. L. Effect of dietary restriction on the fidelity of DNA polymerases in aging mice. *Exp. Gerontol.* **26**, 453–466 (1991).
- Turturro, A. *et al.* Growth curves and survival characteristics of the animals used in the biomarkers of aging program. *J. Gerontol. A Biol. Sci. Med. Sci.* **54**, B492–B501 (1999).
- Yuan, R. *et al.* Genetic coregulation of age of female sexual maturation and lifespan through circulating IGF1 among inbred mouse strains. *Proc. Natl Acad. Sci. USA* **109**, 8224–8229 (2012).
- Zhang, Y. *et al.* The starvation hormone, fibroblast growth factor-21, extends lifespan in mice. *Elife* **1**, e00065 (2012).
- Richardson, A. *et al.* Measures of healthspan as indices of aging in mice—a recommendation. *J. Gerontol. A Biol. Sci. Med. Sci.* <http://dx.doi.org/10.1093/gerona/glv080> (2015).

30. Razzaque, M. S. Does renal ageing affect survival? *Ageing Res. Rev.* **6**, 211–222 (2007).
31. Ferder, L. F., Inserra, F. & Basso, N. Effects of renin-angiotensin system blockade in the aging kidney. *Exp. Gerontol.* **38**, 237–244 (2003).
32. Paul, M., Poyan Mehr, A. & Kreutz, R. Physiology of local renin-angiotensin systems. *Physiol. Rev.* **86**, 747–803 (2006).
33. Bernhard, D. & Laufer, G. The aging cardiomyocyte: a mini-review. *Gerontology* **54**, 24–31 (2008).
34. Sudhir, R., Sukhodub, A., Du, Q., Jovanovic, S. & Jovanovic, A. Ageing-induced decline in physical endurance in mice is associated with decrease in cardiac SUR2A and increase in cardiac susceptibility to metabolic stress: therapeutic prospects for up-regulation of SUR2A. *Biogerontology* **12**, 147–155 (2011).
35. Baker, D. J. *et al.* Increased expression of BubR1 protects against aneuploidy and cancer and extends healthy lifespan. *Nature Cell Biol.* **15**, 96–102 (2012).
36. Childs, B. G., Durik, M., Baker, D. J. & van Deursen, J. M. Cellular senescence in aging and age-related disease: from mechanisms to therapy. *Nature Med.* **21**, 1424–1435 (2015).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank M. Hamada, J. Rainey, Q. Guo, S. Bornschlegel, M. Li, C. M. Roos, N. Hamada and B. Zhang for assistance, and C. Burd, S. Reyes Ramirez, P. Galarzy and D. Katzmman and the members of Program Project Grant AG041122 for discussions. We are grateful to R. Miller and S. Austad for help with the design and interpretation of our lifespan studies. We thank C. Burd

and J. Campisi for sharing *p16-FLUC* and *3MR* MEFs, respectively. This work was supported by the National Institutes of Health (J.M.v.D. R01CA96985 and AG041122 project 1) and (J.D.M., HL111121), the Paul F. Glenn Foundation (J.M.v.D. and D.J.B.), the Ellison Medical Foundation (D.J.B.), the Noaber Foundation (J.M.v.D.), the Children's Research Center (D.J.B.) and the Robert and Arlene Kogod Center on Aging (D.J.B.).

**Author Contributions** D.J.B. performed all lifespan and most healthspan analyses on *ATTAC* mice. B.G.C. designed and conducted experiments to identify and quantify X-gal-positive cells by TEM, and analysed mice for cardiomyocyte hypertrophy and local RAAS activity in kidney. M.E.W., J.Z. and R.A.S. assisted with various aspects of healthspan analyses: the extent of their contributions is reflected in the authorship order. C.J.S. conducted the lifespan analysis of C57BL/6-129Sv/E hybrid mice on diets containing 5% or 9% fat. K.B.J. investigated somatotrophic axis signaling. G.C.C.V., J.D.M. and M.D. analysed resting heart functions by echocardiography. M.D. designed and conducted cardiac stress tests. A.P. and K.K. analysed leukocyte subpopulations. J.M.v.D., D.J.B. and B.G.C. wrote the paper with input from all co-authors. J.M.v.D. directed and supervised all aspects of the study in collaboration with D.J.B.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.M.v.D. ([vandeursen.jan@mayo.edu](mailto:vandeursen.jan@mayo.edu)) or D.J.B. ([baker.darren@mayo.edu](mailto:baker.darren@mayo.edu)).

## METHODS

**Mouse strains and drug treatments.** All studies described using *ATTAC* mice were performed on transgenic line 3, which contains 13 copies of the transgene inserted into a single locus (data available on request). The generation and characterization of the *ATTAC* transgenic line has been previously described in detail<sup>4</sup>. Two cohorts of *ATTAC* mice were consecutively established and analysed. The first cohort consisted of mice on a mixed 129Sv (40% ± 5%) × 57BL/6J (35% ± 5%) × FVB (20% ± 5%) genetic background as determined by Harlan Laboratories using established single nucleotide polymorphism (SNP) panels (Supplementary Table 1). These mice were fed an irradiated pelleted breeder chow diet *ad libitum* for the entire duration of the study: LabDiet product 5058, with 9% fat content (21.635% calories from fat). Mice from this cohort were randomly assigned to be injected with AP20187 (AP; B/B homodimerizer, Clontech) or vehicle twice a week beginning at 12 months of age. Dosing was 0.2 µg AP20187 per g body weight between 12 and 18 months (the same dose as previously used for senescent cell clearance in *BubR1* progeroid mice; see ref. 4), and 2.0 µg AP per g body weight from thereon. The second cohort was established ~1.5 years later after completion of *ATTAC* line 3 backcrossing onto a pure C57BL/6J genetic background (Jackson Lab stock number 000664). Detailed breeding information for the C57BL/6 *ATTAC* cohort is provided in Supplementary Table 2. For the entire duration of the study, congenic *ATTAC* transgenic mice were maintained *ad libitum* on standard irradiated pelleted chow diet: LabDiet product 5053 with 5% fat content (13.205% calories from fat). The number of mice in the C57BL/6J cohort was increased compared to the mixed cohort to allow for more comprehensive healthspan analyses at 18 months of age. The C57BL/6J cohort was randomly assigned to receive either 2.0 µg AP per g body weight or vehicle from 12 months until the end of life. Non-transgenic mice from C57BL/6J *ATTAC* transgenic crosses were used to establish the control cohort for confirmation that AP had no effect on healthspan in the absence of the *ATTAC* transgene. *BubR1*<sup>H/H</sup>; *ATTAC* mice used in this study were generated as previously described<sup>4</sup>. Animals were housed in a pathogen-free barrier environment throughout the duration of the study and kept on a 12-h light–dark cycle. Specifically, mice were housed 5 per cage in static autoclaved HEPA-ventilated microisolator cages measuring 446 cm<sup>2</sup> (27 × 16.5 × 15.5 cm), which are opened only within class II biosafety cabinets. Restricted entry to the room requires key card access to unlock the door. Entrance into the barrier facility requires wearing autoclaved gowns, disposable gloves, caps and shoe covers. Personnel are not allowed to enter any other rodent facility before entry into the barrier. Mouse colonies in this facility are monitored through contact sentinels (one cage per rodent rack), which are evaluated quarterly for (and are free of) the following agents: Sendai virus, pneumonia virus of mice (PVM), mouse hepatitis virus (MHV), minute virus of mice (MVM), Theiler's murine encephalomyelitis virus (TMEV strain GDVII), reovirus, rotavirus (EDIM), mouse parvovirus (MPV strain 1 and 2), murine norovirus (MNV), *Mycoplasma pulmonis*, *Aspiculuris tetraptera* and *Syphacia* spp. In addition, the following are annually screened for (and free of) these agents: lymphocytic choriomeningitis virus (LCMV), ectromelia virus (mousepox), K virus, polyoma virus, mouse adenovirus (MAV type 1 and 2), hantavirus, Prospect Hill virus (PHV), mouse cytomegalovirus (MCMV), lactate dehydrogenase elevating virus (LDEV) *Encephalitozoon cuniculi*, cilia-associated respiratory bacillus (CARB), mouse thymic virus (MTV, MTLV), *Clostridium piliforme*, *Myobia* spp. and *Myocoptes* spp. Autoclaved Enrich-o-Cobs (The Andersons Incorporated) were used as bedding. Cages and bedding were changed weekly. Mice had *ad libitum* access to municipal city water that was sterilized with ultraviolet light. Room temperature was maintained between 21 °C and 23 °C. Ear tags (1005-1, National Band and Tag Co.) were applied to pups at the time of genotyping for identification purposes. All animal procedures were approved by the Mayo Clinic Institutional Animal Care and Use Committee.

**Statistical analysis.** Prism software was used for statistical analysis and generation of survival, cataract and tumour latency curves. Two-tailed unpaired *t*-tests were used for pairwise significance unless otherwise indicated. log-rank tests were used for survival, cataract and tumour latency curves. For maximum lifespan extension analysis, two-sided Wang–Allison tests were used<sup>14</sup>. For consistency in these comparisons, the following denotes significance in all figures: \**P* < 0.05, \*\**P* < 0.01 and \*\*\**P* < 0.001. We note that no power calculations were used. Sample sizes were based on previously published experiments where differences were observed. No samples were excluded. Investigators were blinded to allocation during experiments and outcome assessment, except for rare instances where blinding was not possible.

**Single-cell preparations and flow cytometry.** To determine the cell type origin of senescent cells accumulating in fat, we prepared single cells from iWAT, collected the stromal vascular fraction by centrifugation and labelled the resulting cells with specific cell surface markers for collection of the CD45<sup>+</sup>, endothelial, and progenitor (adipocyte stem cells plus preadipocytes) cell populations by FACS as previously described<sup>4,35,37</sup>. To quantify progenitor cells, we divided the number of progenitors

recovered from the stromal vascular fraction by the total number of cells in the stromal vascular fraction. Single cell preparations from wounds were prepared in a similar manner. The proportion of GFP<sup>+</sup> cells in each population was assessed using flow cytometry as described<sup>4</sup>.

**Analysis of age-related phenotypes.** Bi-weekly checks for cataracts were performed as described<sup>5</sup>. iWAT cell diameter, adipose depot weight, skin thickness, and skeletal muscle diameter measurements were as described<sup>4</sup>. Open field-testing was performed in a 50 cm × 50 cm box and analysed with TopScan software (CleverSys Inc.). Sclerotic glomeruli and blood urea nitrogen assessments were performed as described<sup>38</sup>. To determine cardiomyocyte cross-sectional area, three sagittal sections of the mid-ventricle separated by 400 µm were stained with periodic acid-Schiff using a kit (Poly Scientific R&D, k098). At least 150 cardiomyocytes in the left ventricular free wall per section were measured using Image J software. Lethal isoproterenol challenges were performed as described<sup>38</sup>. Parameters of cardiac function were determined with echocardiography using Vevo2100 Imaging System (VisualSonics Inc.) with a 30 MHz transducer. Mice were lightly anaesthetized by isoflurane (0.8–1.5%) with minimal effect on heart rate. Several m-mode recordings of the cardiac parasternal short axis at the level of papillary muscles were made and subsequently analysed. For non-lethal isoproterenol challenge, mice were analysed at baseline (day 0) and subsequently were administered five consecutive days of treatment with 10 mg kg<sup>−1</sup> twice daily as described<sup>39</sup>. Mice were then assessed for alteration in LV mass on day 6. For longitudinal measurements of cardiac structure recordings were acquired in conscious animals. At least three different recordings per animal at each time point were analysed. Rotarod assessments were as previously published<sup>40</sup>. Exercise ability was performed as described<sup>4</sup>. Novel object discrimination tests were performed as described<sup>41</sup>. Grip strength was determined as previously published<sup>42</sup>. Gross haematology was assessed by analysing EDTA-treated whole blood using a Hemavet 950 (Drew Scientific Inc.). FACS-based analyses of peripheral blood leukocyte populations were as previously described<sup>43</sup>. CD3<sup>+</sup> T lymphocytes were collected from peripheral blood (isolated using CD3 MicroBeads and MACS columns purchased from Miltenyi Biotec) according to manufacturer's instructions. Glucose tolerance was assessed in mice fasted overnight by measuring blood glucose levels before and at time points after an intraperitoneal injection of glucose (1.0 g kg<sup>−1</sup>). Insulin sensitivity was measured following a 4 h fast by measuring blood glucose levels before and at time points after an intraperitoneal injection of insulin (0.6 mU g<sup>−1</sup>). Serum Igf1 concentrations were measured by ELISA according to manufacturer's instructions (R&D Systems, MG100). Wound healing assessments were performed as previously described<sup>44</sup>. Phosphotungstic acid haematoxylin (PTAH) staining on formalin-fixed, paraffin-embedded tissue samples was performed as described previously<sup>35</sup>.

**Quantitative PCR.** RNA extraction, cDNA synthesis and qRT-PCR analysis of various mouse tissues and CD3<sup>+</sup> T lymphocytes collected from peripheral blood were performed as described<sup>5</sup>. RNA extraction from cell lines was performed according to manufacturer's instructions (Qiagen RNeasy mini kit). Sequences of primers for *FKBP-Casp8*, *Egfp*, *Ink4a*, *Arf*, *p21*, *Il6*, *Pai1*, *Igf1*, *p2* and *Mmp13* were as previously published<sup>4</sup>. In addition, the following primers were used: *Mmp3* forward 5'-CCTGATGT-TGGTGCTTCA-3', reverse 5'-TCCTGTAGGTGATGTGGGATTTC-3'; *Pparg* forward 5'-TCTTCCATCACGGAGAGGTC-3', reverse 5'-GATGCACTGCCTATGAGC-AC-3'; *Cebpa* forward 5'-TTGTTTGGCTTTATCTCGGC-3', reverse 5'-CCAAGAAG-TCGGTGACAAG-3'; *Il1a* forward 5'-TCAACCAACTATATATATCAGGATGTGG-3', reverse 5'-CGAGTAGGCATACATGTCAAATTTTAC-3'; *Tnfa* forward 5'-AGGGTCTGGGCCATAGAACT-3', reverse 5'-CAGCCTCTTCTCATTCCTGC-3'; *Agtr1a* forward 5'-AAGGGCCATTTTGCTTT-TCT-3', reverse 5'-AACTCAGCAACCCTCCAA-3'; *Sur2a* forward 5'-CAATGAGGCTGTGACAGACA-3', reverse 5'-GGAGCTGACAGACACGAACA-3'; and *Igf1bp5* forward 5'-ATACAACCCAGACGCCAGCT-3', reverse 5'-ACCTGGGCTATGCACTTGATG-3'.

**SA-β-Gal staining, electron microscopy and X-ray elemental analysis.** For SA-β-Gal staining on kidney cryosections, tissue was embedded in OCT, sectioned at 10 µm thickness, and air-dried. After rehydration in PBS, staining was performed using the Cell Signaling kit (9860) with a 12-min fixation followed by incubating in the staining solution at 37 °C for 12 h. Whole mount SA-β-Gal stainings on fat, kidney and heart were performed using a kit (Cell Signaling), with a 20-min fixation followed by immersing samples in the staining solution at 37 °C for 12 h (iWAT and eWAT) or 48 h (heart and kidney). Detection of X-Gal crystals by transmission electron microscopy was as previously described<sup>45</sup>. Specifically, X-Gal-containing whole mounts were fixed in Trump's fixative at 4 °C for 12 h and routinely processed for OsO<sub>4</sub>/lead staining. Images were acquired and quantification performed on a Jeol 1400+ electron microscope with 80 kV acceleration voltage. The authenticity of X-Gal crystals was confirmed by X-ray elemental analysis, whereas the specificity of crystal formation was verified using fat and heart tissue that was not



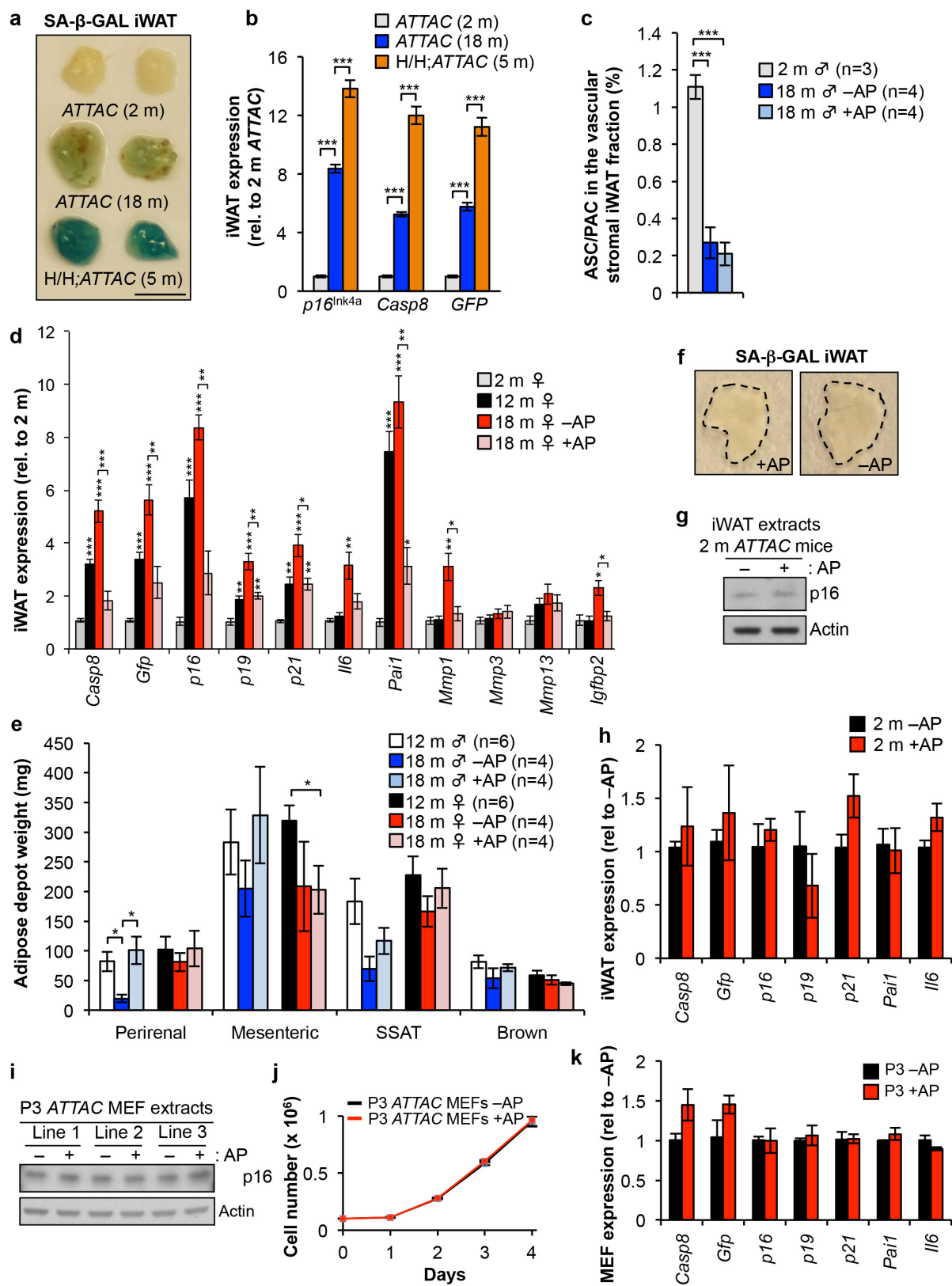
stained for SA- $\beta$ -Gal. To measure the percentage of X-Gal-crystal-containing cells in kidney sections of AP- or vehicle-treated mice, we screened 5 grids ( $>1,500$  cells per grid) per treatment group for cells X-gal-positive cells at  $3,000\times$  magnification. Cells with one or more crystals and the total number of cells were counted. iWAT and the visceral pericardium were assessed similarly, with cells with one or more crystals considered to be X-Gal positive. For iWAT, we only observed crystals in stromovascular cells with morphology and perivascular location consistent with adipocyte progenitors and therefore expressed the X-Gal fraction of total cells. More than 500 total fat cells per animal were scored for iWAT ( $n = 4$  animals per group). Total cells was estimated by extrapolating from the cellularity of five  $1,500\times$  fields. Owing to dissection and processing, the serous pericardium was lost for most samples, so the visceral pericardium alone was scored via TEM. Only ciliated epithelial cells stained positive for SA- $\beta$ -Gal, therefore we expressed the X-Gal positivity of this tissue as percentage of this cell type. The entire visible pericardium on one TEM grid per animal was scored, with four animals per group. For quantitation of SA- $\beta$ -Gal positive cells in GFP<sup>+</sup> and GFP<sup>-</sup> cell fractions collected from iWAT, 100 cells were scored per fraction ( $n = 3$  animals).

**Tissue immunofluorescence staining and western blotting.** Paraffin-embedded kidney tissue sections ( $5\mu\text{m}$ ) were rehydrated ( $2\times 10$  min xylene;  $2\times 5$  min 100% ethanol;  $1\times 5$  min 95% ethanol;  $1\times 5$  min 70% ethanol), incubated in protease from *Streptomyces griseus* (Sigma, P6911;  $0.75\text{ mg ml}^{-1}$  in Tris buffer, 0.05 M, pH 7.6) for 1 h at  $37^\circ\text{C}$  to retrieve antigens, blocked with Rodent Block M (BioCare Medical) for 30 min, and incubated with rabbit anti-Agtr1a antibody (Novus Bio, NBP1-77078; 1:250) in antibody diluent (Dako) for 12 h at  $4^\circ\text{C}$ . Western blotting on tissue lysates was performed according to standard methods, incubating the nitrocellulose membrane overnight at  $4^\circ\text{C}$  with the following antibodies: Agtr1a (as above; 1:750), S6K (Cell Signaling, 2708; 1:1,000), phospho-S6K (Cell Signaling, 9234; 1:1,000), AKT (Cell Signaling, 9272; 1:1,000), and phospho-AKT<sup>S473</sup> (Cell Signaling, 9271; 1:1,000). Quantification of p-S6K/S6K and p-AKT<sup>S473</sup>/AKT ratios was performed as described<sup>46,47</sup>.

**Cell culture.** ATTAC MEFs were generated as previously described<sup>4</sup>. Western blotting for p16<sup>INK4a</sup> and mouse anti-actin was performed as described<sup>48</sup>. Growth curves were generated as previously described<sup>48</sup>. RNA extraction was done with a Micro RNeasy kit according to manufacturer's instructions (Qiagen). Early passage p16-3MR MEFs<sup>8</sup> were provided by J. Campisi and M. Demaria. Early passage

homozygous p16-FLUC MEFs<sup>49</sup> were provided by C. Burd. MEFs were immortalized by expression of SV40 large T antigen as described<sup>50</sup>. The cell line genotypes were authenticated by PCR genotyping. They were not tested for mycoplasma contamination.

37. Baker, D. J., Weaver, R. L. & van Deursen, J. M. p21 both attenuates and drives senescence and aging in BubR1 progeroid mice. *Cell Rep.* **3**, 1164–1174 (2013).
38. Baker, D. J. *et al.* Increased expression of BubR1 protects against aneuploidy and cancer and extends healthy lifespan. *Nature Cell Biol.* **15**, 96–102 (2013).
39. Zingman, L. V. *et al.* Kir6.2 is required for adaptation to stress. *Proc. Natl Acad. Sci. USA* **99**, 13278–13283 (2002).
40. Wijshake, T. *et al.* Reduced life- and healthspan in mice carrying a mono-allelic *BubR1* MVA mutation. *PLoS Genet.* **8**, e1003138 (2012).
41. Buenz, E. J. *et al.* Apoptosis of hippocampal pyramidal neurons is virus independent in a mouse model of acute neurovirulent picornavirus infection. *Am. J. Pathol.* **175**, 668–684 (2009).
42. Maskey, R. S. *et al.* Spartan deficiency causes genomic instability and progeroid phenotypes. *Nature Commun.* **5**, 5744 (2014).
43. Blatner, N. R. *et al.* Expression of ROR- $\gamma$ t marks a pathogenic regulatory T cell subset in human colon cancer. *Sci. Transl. Med.* **4**, 164ra159 (2012).
44. Baker, D. J. *et al.* BubR1 insufficiency causes early onset of aging-associated phenotypes and infertility in mice. *Nature Genet.* **36**, 744–749 (2004).
45. Stollewerk, A., Klambt, C. & Cantera, R. Electron microscopic analysis of *Drosophila* midline glia during embryogenesis and larval development using  $\beta$ -galactosidase expression as endogenous cell marker. *Microsc. Res. Tech.* **35**, 294–306 (1996).
46. Garelick, M. G. *et al.* Chronic rapamycin treatment or lack of S6K1 does not reduce ribosome activity *in vivo*. *Cell Cycle* **12**, 2493–2504 (2013).
47. Schreiber, K. H. *et al.* Rapamycin-mediated mTORC2 inhibition is determined by the relative expression of FK506-binding proteins. *Aging Cell* **14**, 265–273 (2015).
48. Baker, D. J. *et al.* BubR1 insufficiency causes early onset of aging-associated phenotypes and infertility in mice. *Nature Genet.* **36**, 744–749 (2004).
49. Burd, C. E. *et al.* Monitoring tumorigenesis and senescence *in vivo* with a p16<sup>INK4a</sup>-Luciferase model. *Cell* **152**, 340–351 (2013).
50. Ricke, R. M., Jeganathan, K. B., Malureanu, L., Harrison, A. M. & van Deursen, J. M. Bub1 kinase activity drives error correction and mitotic checkpoint control but not tumor suppression. *J. Cell Biol.* **199**, 931–949 (2012).

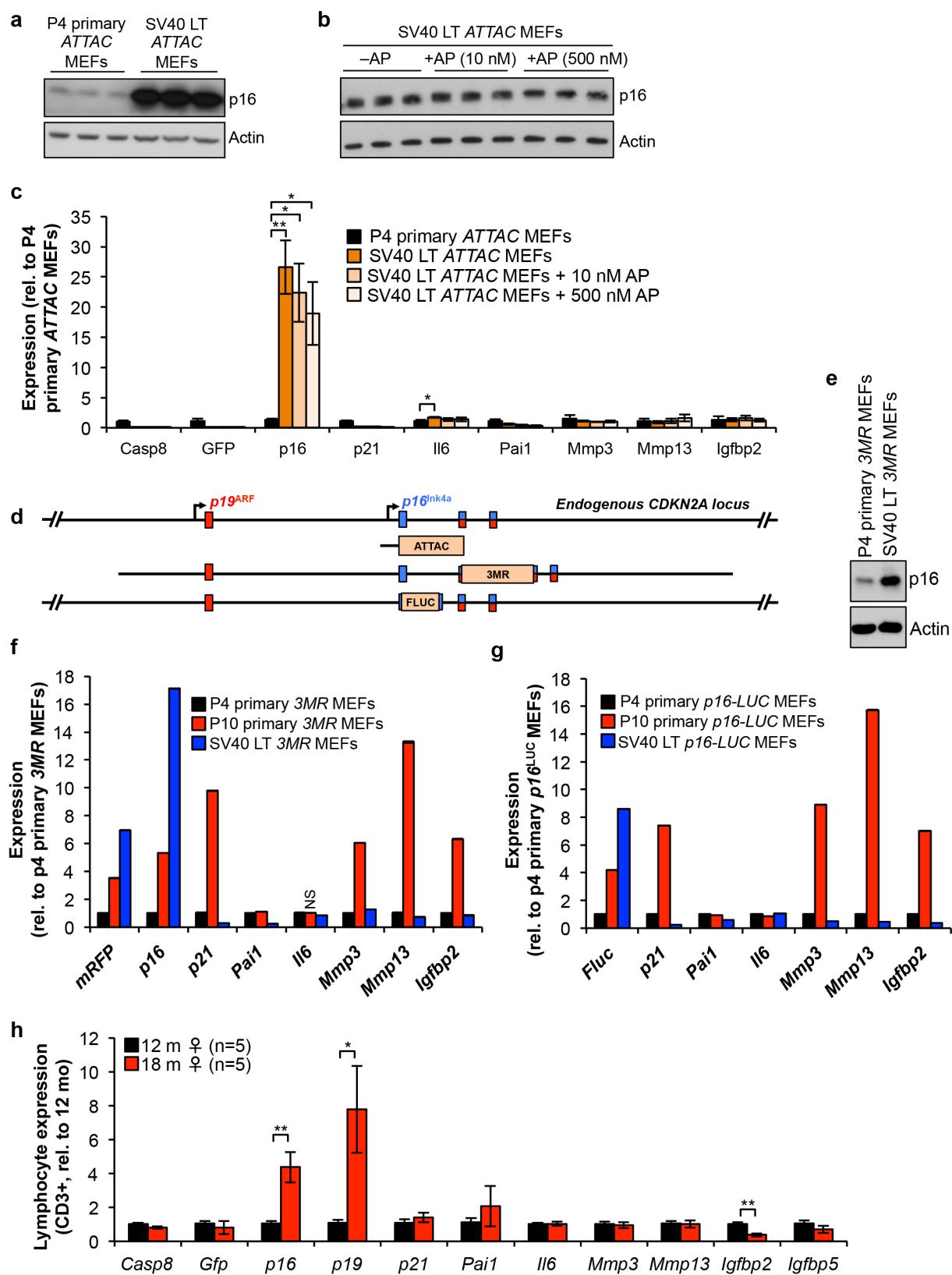


Extended Data Figure 1 | See next page for figure caption.

**Extended Data Figure 1 | *ATTAC* transgene expression tracks with expression of senescence markers in iWAT and induces apoptosis of senescent cells after AP administration.** **a**, Comparative analysis of SA- $\beta$ -Gal activity in intact iWAT. Scale bar, 0.5 cm. **b**, Analysis of endogenous *Ink4a* and *ATTAC* transcript SA- $\beta$ -Gal activity in iWAT by qRT-PCR. H/H denotes *BubR1*<sup>H/H</sup> mice ( $n = 4$  mice per group). **c**, FACS-based quantification of iWAT progenitor cell numbers in 18-month-old *ATTAC* mice treated with vehicle or AP. ASC, adipocyte stem cells; PAC, preadipocytes. **d**, Expression of the *ATTAC* transgene and senescence markers in iWAT as determined by qRT-PCR ( $n = 4$  mice per group). Asterisks above individual bars denote significant changes to 2-month-old mice; asterisks above brackets denote significant differences between 18-month-old vehicle and AP-treated mice. **e**, Perirenal, mesenteric, subscapular and brown adipose tissue depot weights. SSAT, subscapular adipose tissue. **f**, SA- $\beta$ -Gal activity in iWAT from 2-month-old *ATTAC*

mice treated with vehicle or AP beginning at weaning age. **g**, p16<sup>Ink4a</sup> levels in iWAT from the mice described in **f**. Actin was used as a loading control. **h**, Expression of *ATTAC* and senescence marker mRNA in the mice described in **f** ( $n = 3$  mice per group). **i–k**, Early passage non-senescent *ATTAC* MEFs express p16<sup>Ink4a</sup> but are not susceptible to FKBP–Casp8-mediated elimination when cultured in the presence of AP. **i**, Levels of p16<sup>Ink4a</sup> in passage 3 *ATTAC* MEFs, with and without AP treatment. **j**, Growth curves of passage 3 *ATTAC* MEFs ( $n = 4$  independently generated MEF lines per group), with or without AP treatment. **k**, Expression of *ATTAC* and senescence marker mRNA in passage 3 *ATTAC* MEFs ( $n = 3$  independently generated MEF lines per group), with or without AP treatment. Error bars indicate s.e.m. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  (unpaired two-tailed  $t$ -tests). For gel source data, see Supplementary Fig. 1.

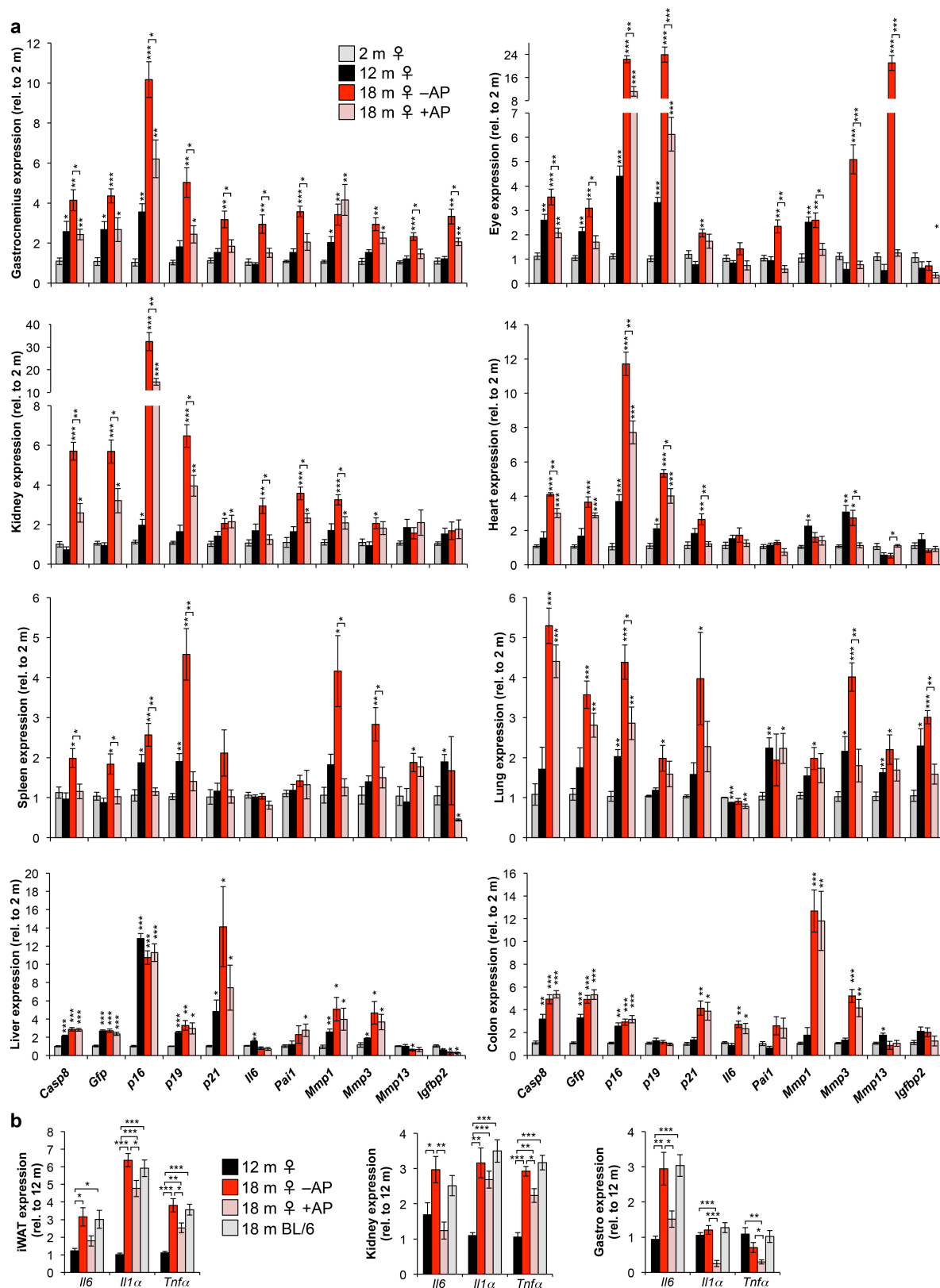




Extended Data Figure 2 | See next page for figure caption.

**Extended Data Figure 2 | *ATTAC* lacks promoter elements required for expression in replication-competent cells or aged lymphocytes expressing high levels of endogenous p16<sup>Ink4a</sup>.** **a–c**, SV40 large-T-antigen-immortalized *ATTAC* MEFs robustly express endogenous p16<sup>Ink4a</sup> (owing to SV40 large-T-antigen-mediated inactivation of Rb) but fail to engage the *ATTAC* transgene and are not subject to FKBP–Casp8-mediated elimination. **a**, p16<sup>Ink4a</sup> protein levels in passage 4 (P4) primary MEFs and MEFs immortalized with SV40 large T antigen. Actin was used as a loading control. **b**, p16<sup>Ink4a</sup> protein levels in immortalized MEFs treated with vehicle or two concentrations of AP. Actin was used as a loading control. **c**, Expression of *ATTAC* and senescence marker transcripts in passage 4 primary MEFs, vehicle-treated immortalized MEFs, and AP-treated immortalized MEFs ( $n = 3$  independently generated MEF lines per group). **d**, Schematic representation of the endogenous *Ink4a* locus and the various *Ink4a* promoter regions driving *ATTAC*, *3MR* and firefly luciferase (*FLUC*). *ATTAC* and *p16-3MR* mice have 2.6 kb and ~50 kb *Ink4a* promoter fragments driving transgene activity, respectively.

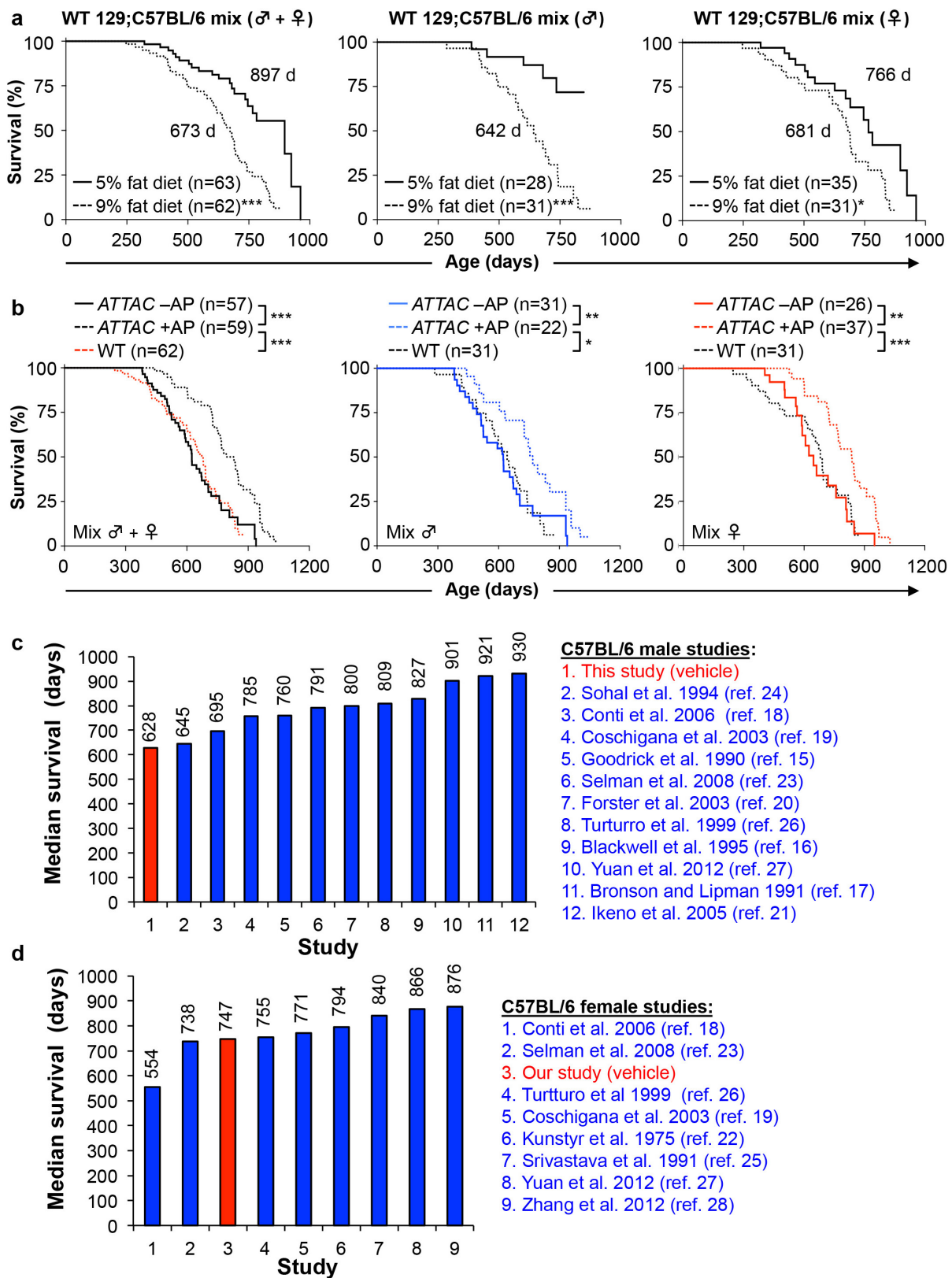
*p16-FLUC* has firefly luciferase knocked into the endogenous *Ink4a* locus, which keeps the entire promoter region intact but ablates p16<sup>Ink4a</sup> protein expression. **e**, p16<sup>Ink4a</sup> protein levels in early passage primary and SV40 large-T-antigen-immortalized *p16-3MR* MEFs. **f**, Expression of senescence marker mRNA in early and late passage primary MEFs and SV40 large-T-antigen-immortalized *p16-3MR* MEFs ( $n = 1$  independently generated MEF line per group performed in triplicate). **g**, Expression of senescence marker mRNA in early and late passage primary MEFs and SV40 large-T-antigen-immortalized *p16-FLUC* MEFs ( $n = 1$  independently generated MEF line per group performed in triplicate). **h**, Expression of *ATTAC* and senescence markers in CD3<sup>+</sup> T cells from 12- and 18-month-old *ATTAC* mice ( $n = 5$  mice per group). Error bars indicate s.e.m. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  (unpaired two-tailed  $t$ -tests). We note that all values in **f** and **g** have  $P < 0.05$  compared to passage 4 MEFs, with the exception of the one marked NS for not significant. For gel source data, see Supplementary Fig. 1.



**Extended Data Figure 3 | ATTAC-mediated clearance of senescent cells is partial and tissue selective and attenuates expression of inflammation markers.** **a**, Expression of the ATTAC transgene and a senescence marker panel, as determined by RT-PCR, in gastrocnemius, eye, kidney, heart (atria), spleen, lung, liver and colon ( $n = 4$  females per group). **b**, Expression of *Il6*, *Il1a* and *Tnfa* as determined by qRT-PCR in mouse iWAT, kidney and skeletal muscle at different ages ( $n = 4$  females per group). *Il6* values are as indicated in Extended Data Fig. 1d (iWAT) and

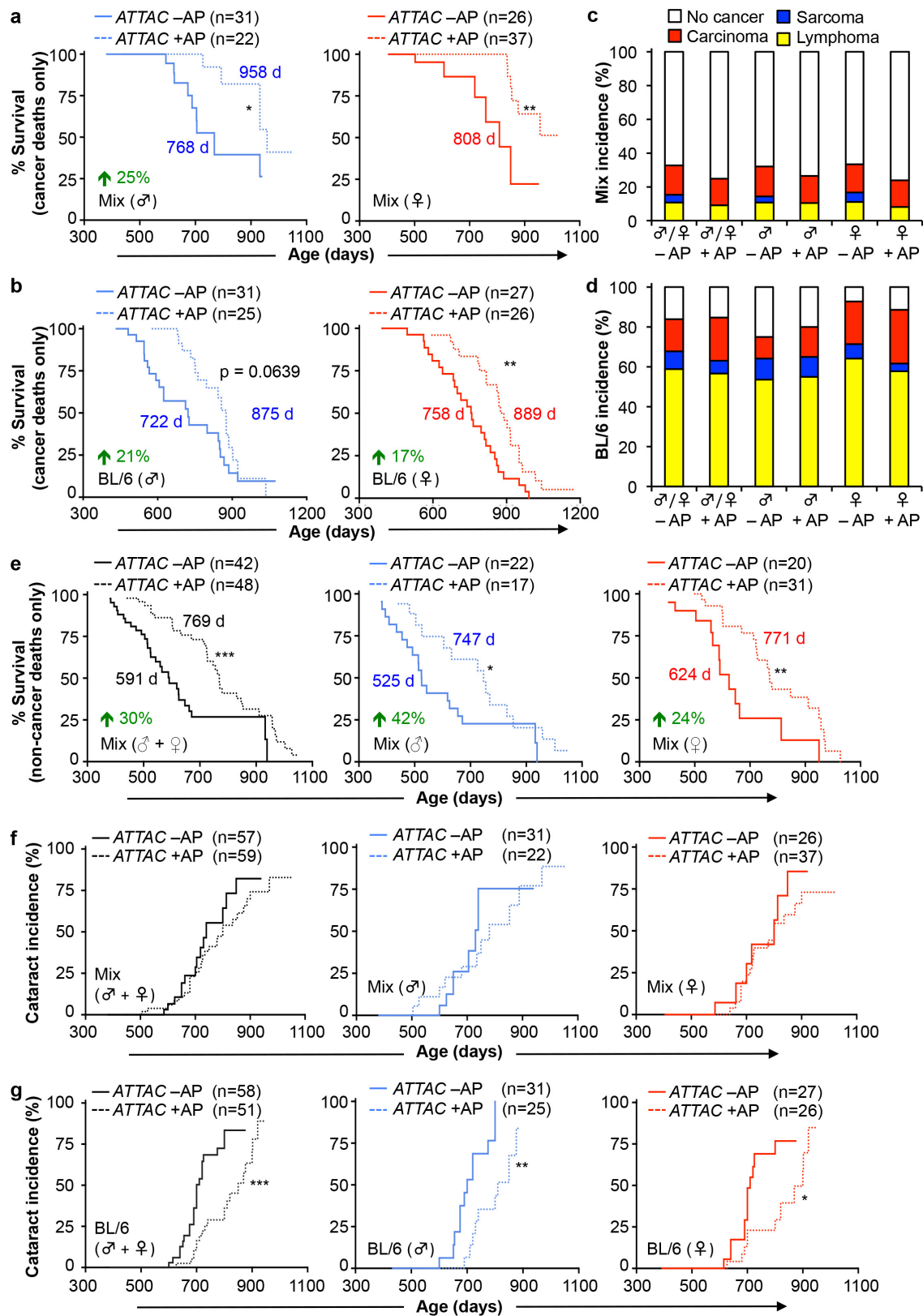
in **a** (kidney and gastrocnemius). Expression levels of inflammation markers in unmanipulated 18-month-old C57BL/6 females suggests that repeated vehicle injections were not a source of tissue inflammation. Error bars indicate s.e.m. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  (unpaired two-tailed  $t$ -tests). Asterisks above individual bars in **a** denote significant changes to 2-month-old mice; asterisks above brackets denote significant differences between 18-month-old vehicle and AP-treated mice.





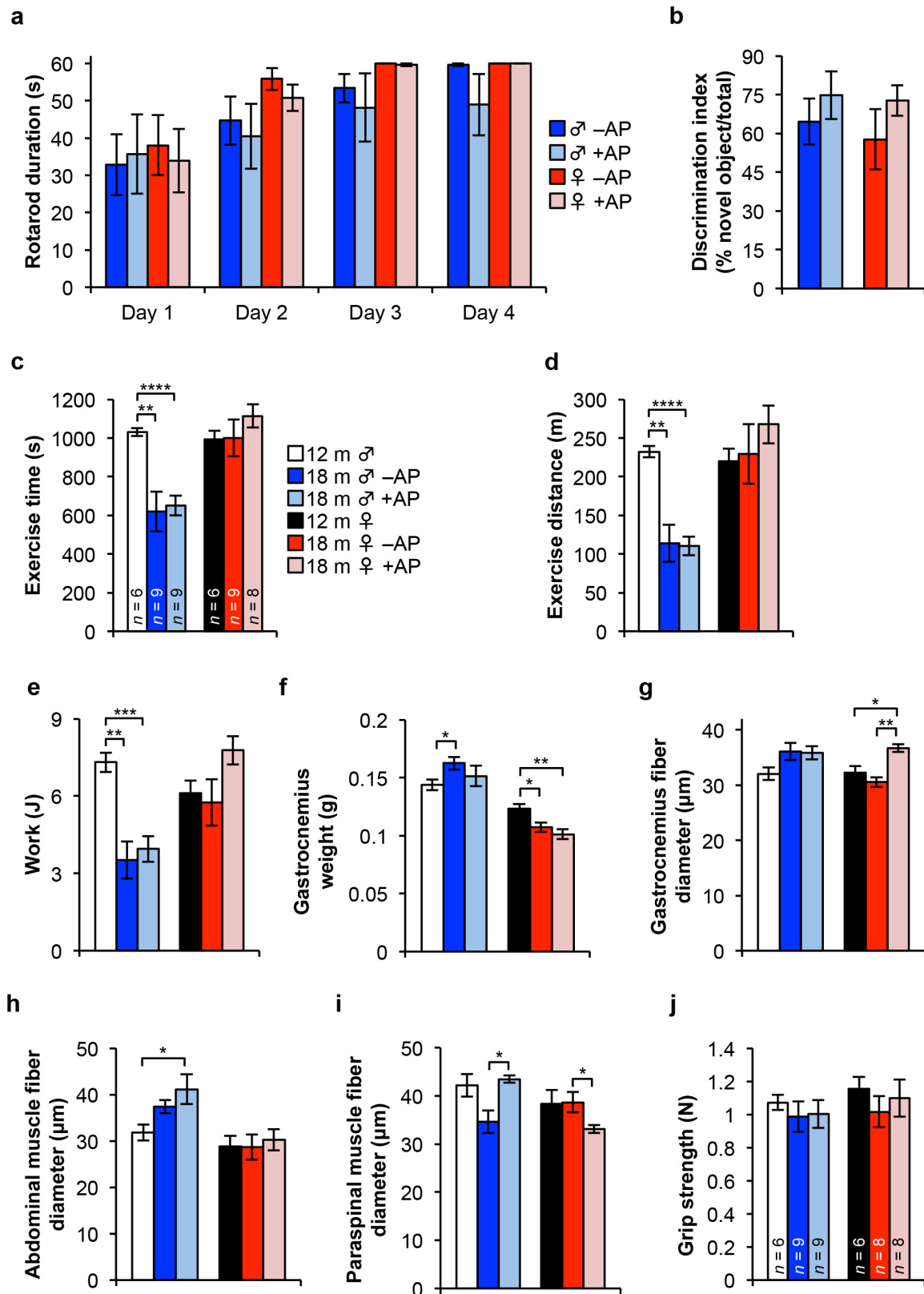
**Extended Data Figure 4 | Comparison of lifespans under different diets and housing facilities.** **a**, Survival curves of unmanipulated wild-type C57BL/6-129Sv mice fed a 5% versus 9% fat diet. Median lifespan (in days) are indicated. **b**, Survival curves of unmanipulated wild-type C57BL/6-129Sv mice fed a 9% fat diet plotted against those of vehicle-treated (-AP) and AP-treated (+AP) C57BL/6-129Sv-FVB ATTAC mice from Fig. 2b.

These data suggest that the lifespans of vehicle-injected C57BL/6-129Sv-FVB control mice were quite normal for the diet that they were on, and unlikely to be negatively affected by repeated intraperitoneal injections. \* $P < 0.01$ ; \*\* $P < 0.001$ ; \*\*\* $P < 0.001$  (log-rank tests). **c**, **d**, Median survival data of unmanipulated C57BL/6 male (**c**) and female (**d**) mice from various laboratories for comparison to the results obtained from our facility.



**Extended Data Figure 5 | Senescent cell clearance delays tumour and cataract formation.** **a, b**, Survival curves of mixed (a) and C57BL/6 (b) ATTAC mice dying of cancer (mice that had an overt tumour at time of death; only mice with lymphomas, sarcomas and carcinomas were included). Median survival (in days) and percentage increase are indicated. **c, d**, Incidence of macroscopically detectable neoplasms (lymphomas, sarcomas and carcinomas) at time of death in mixed (c) and C57BL/6

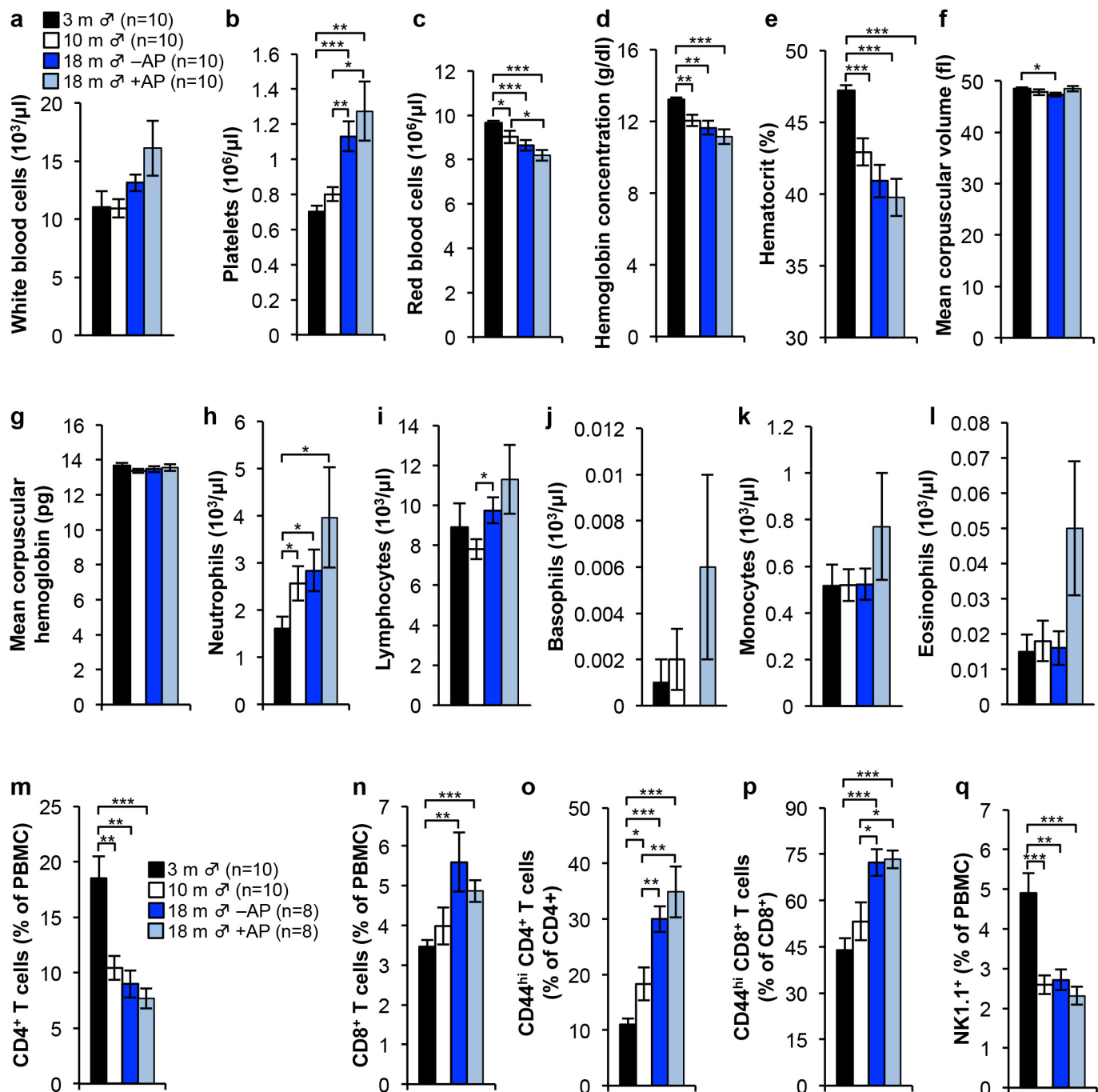
(d) ATTAC mice from survival cohorts. **e**, Survival curves of C57BL/6-129Sv-FVB mice dying without cancer (mice that had an overt tumour at time of death, including lymphoma, sarcoma and/or carcinoma, were excluded). Median survival (in days) and percentage increase are indicated. **f, g**, Cataract incidence for mixed (f) and C57BL/6 (g) ATTAC mouse cohorts. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  (log-rank tests).



**Extended Data Figure 6 | Senescent cell clearance does not affect coordination, memory or exercise ability of 18-month-old ATTAC mice.** **a**, Time spent balanced during a fixed speed rotarod test for 18-month-old ATTAC mice ( $n = 6$  male and 8 female mice per group). **b**, Novel object investigation test. The percentage of investigations of a novel object divided by the total investigations is graphed. Key and animal numbers are as indicated in **a**. **c–e**, Time-to-exhaustion (**c**), distance (**d**) and work (**e**) during a treadmill exercise test. Animal numbers are as

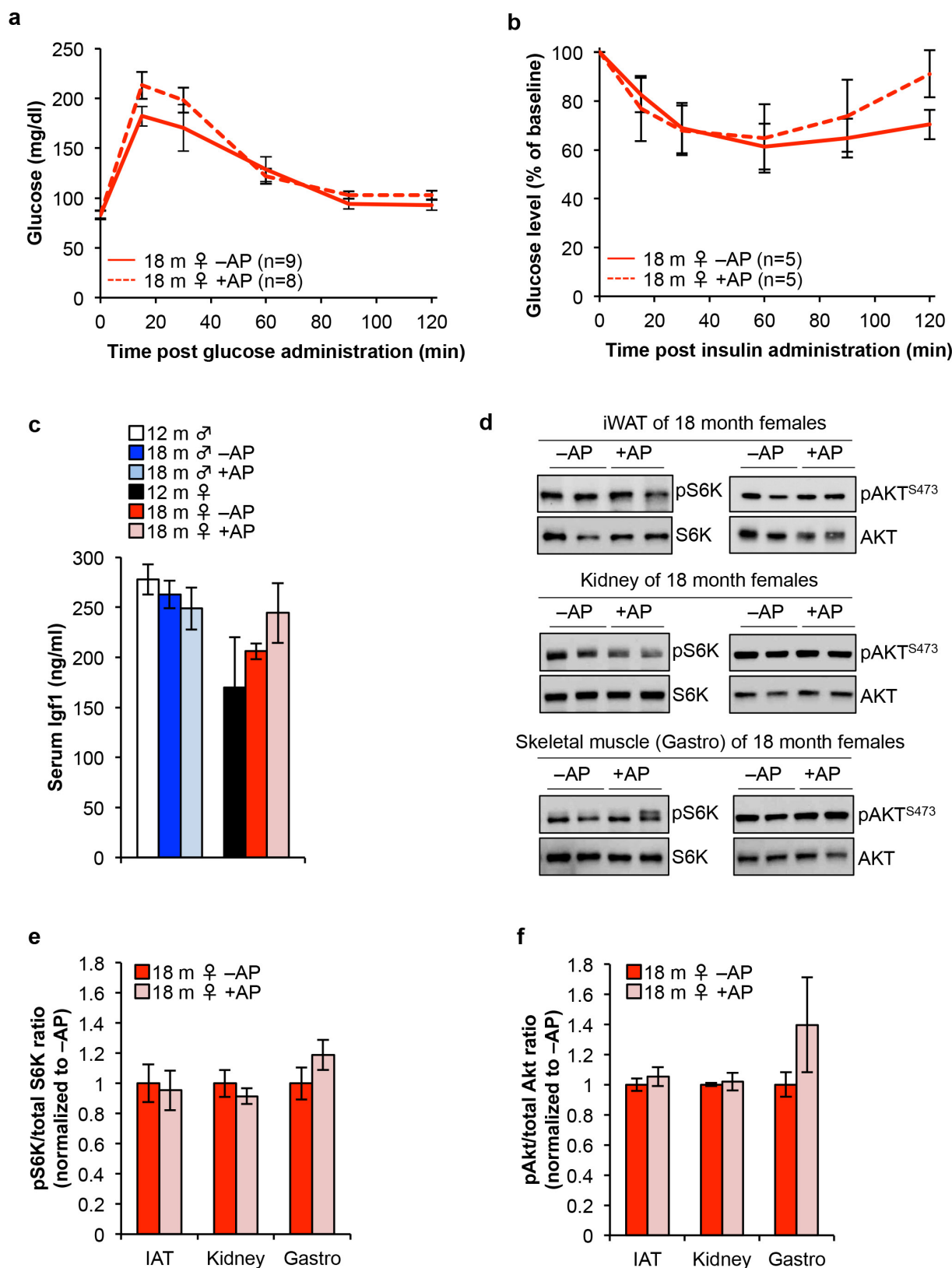
indicated in **c**, **f**, Gastrocnemius muscle weight of ATTAC mice ( $n = 6$  12-month-old males and females;  $n = 4$  18-month-old –AP males and females;  $n = 4$  18-month-old +AP males and females). **g–i**, Myofibre diameter measurements on isolated gastrocnemius (**g**), abdominal (**h**) and paraspinal muscle (**i**). Animal numbers are as indicated in **f**, **j**, Analysis of forelimb grip strength of ATTAC mice. Error bars indicate s.e.m. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  (unpaired two-tailed  $t$ -tests).





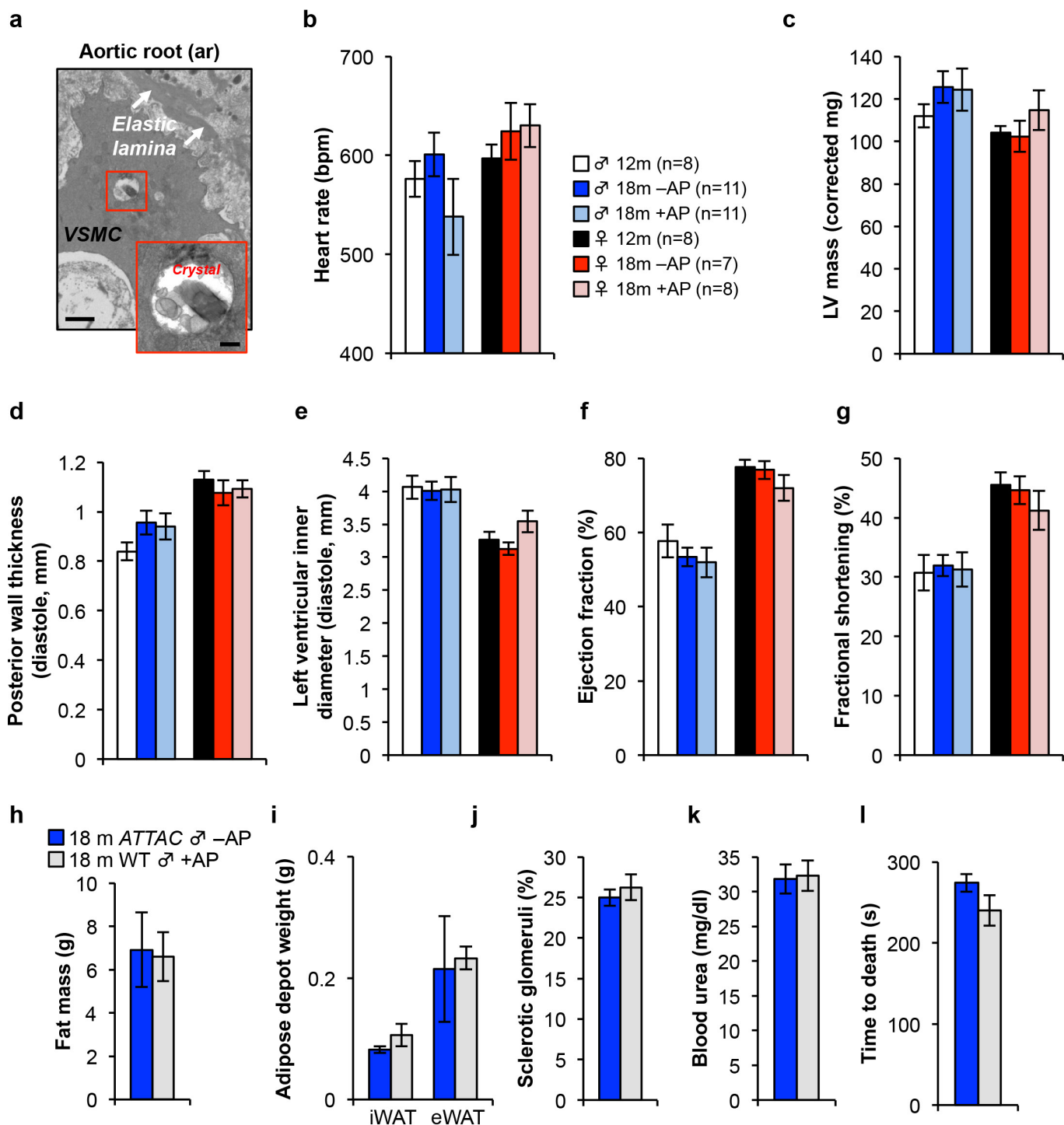
**Extended Data Figure 7 | Senescent cell clearance has no effect on haematological parameters and age-related changes in leukocyte populations.** a–l, Haematology results of 3- and 10-month-old untreated ATTAC C57BL/6 mice and 18-month-old vehicle- and AP-treated ATTAC C57BL/6 mice. White blood cell count (a), platelet count (b), red blood cell count (c), haemoglobin concentration (d), haematocrit (e), mean corpuscular volume (f), mean corpuscular haemoglobin (g), neutrophils (h), lymphocytes (i), basophils (j), monocytes (k) and eosinophils (l).

m–q, Assessment for leukocyte subpopulations in 3- and 10-month-old untreated ATTAC C57BL/6 mice and 18-month-old vehicle- and AP-treated ATTAC C57BL/6 mice. CD4<sup>+</sup> T cells (percentage of peripheral blood mononuclear cells, PBMC) (m), CD8<sup>+</sup> T cells (percentage of PBMC) (n), CD44<sup>hi</sup> CD4<sup>+</sup> T cells (percentage of CD4<sup>+</sup>) (o), CD44<sup>hi</sup> CD8<sup>+</sup> T cells (percentage of CD8<sup>+</sup>) (p), and NK1.1<sup>+</sup> cells (percentage of PBMC) (q). Error bars indicate s.e.m. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  (unpaired two-tailed  $t$ -tests).



**Extended Data Figure 8 | Senescent cell removal does not affect somatotrophic axis signalling *in vivo*.** **a**, Glucose levels following intraperitoneal glucose administration after an overnight fast in 18-month-old vehicle- and AP-treated *ATTAC* C57BL/6 females. **b**, Normalized glucose levels after intraperitoneal insulin administration following a 4-h fast in 18-month-old vehicle- and AP-treated *ATTAC* C57BL/6 females. **c**, Serum Igf1 levels in *ATTAC* C57BL/6 mice ( $n = 4$  mice of each group). **d**, Representative western blots for phospho-S6K,

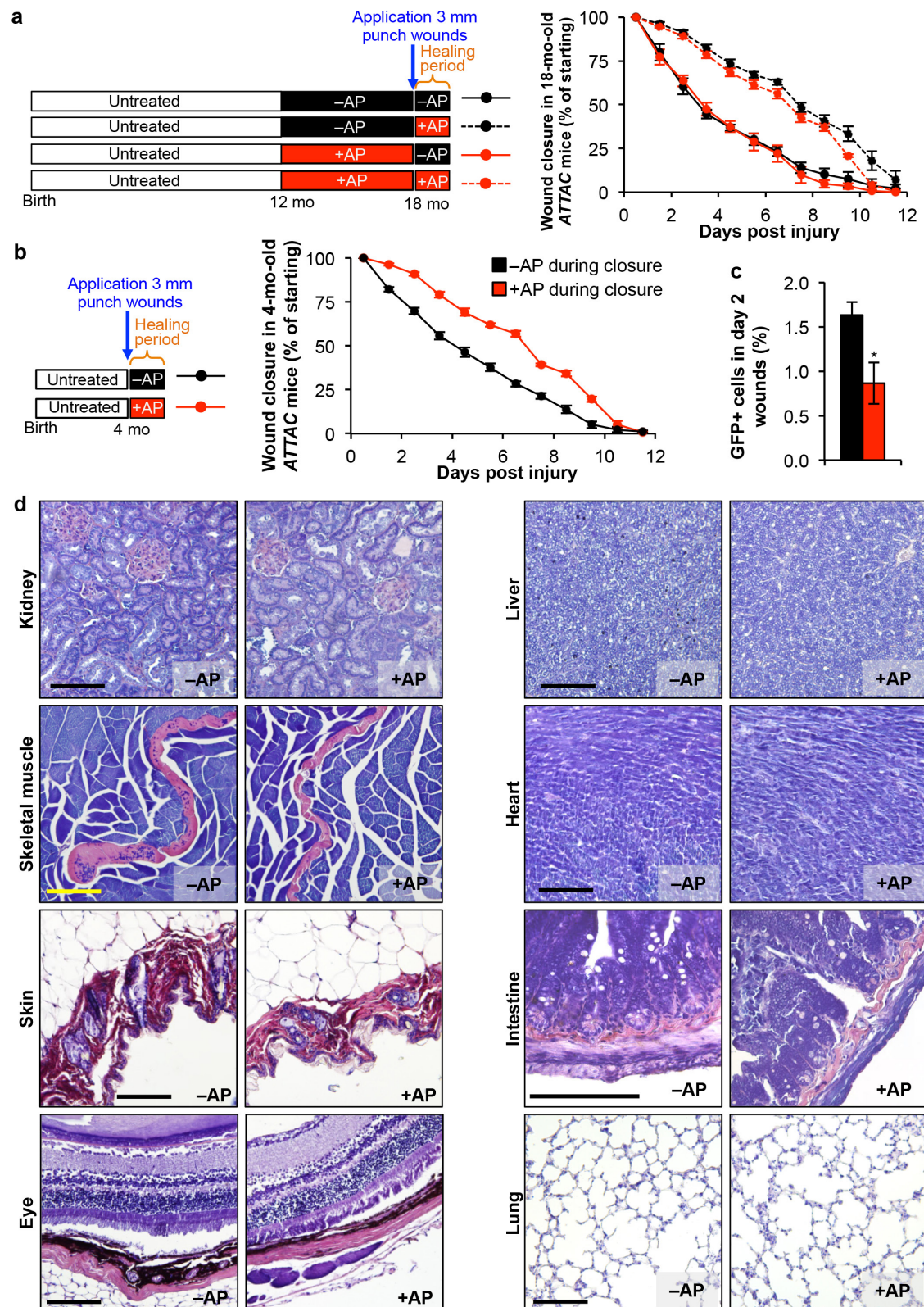
total S6K, phospho-AKT<sup>S473</sup> and total AKT in iWAT, kidney and skeletal muscle tissue lysates from 18-month-old vehicle- and AP-treated *ATTAC* C57BL/6 females. **e**, Quantification of phospho-S6K to total S6K ratio in blots from **d**,  $n = 4$  mice of each group. **f**, Quantification of phospho-AKT<sup>S473</sup> to total AKT ratio in blots from **d**,  $n = 4$  mice of each group. Error bars indicate s.e.m. No statistically significant differences were observed in **a–c**, **e** and **f** using unpaired two-tailed *t*-tests. For gel source data, see Supplementary Fig. 1.



**Extended Data Figure 9 | Senescent cell clearance does not alter cardiac morphology and function in 'resting' mice and AP treatment has no effect on healthspan of mice lacking the ATTAC transgene.** **a**, Electron micrographs of X-Gal crystal containing cells in the aortic root. VSMC, vascular smooth muscle cell. Scale bars, 1  $\mu$ m (main panel) and 200 nm (inset). **b–g**, Echocardiography measurements of heart rate (**b**), left ventricular mass (**c**), posterior wall thickness (**d**), left ventricular inner diameter (**e**), ejection fraction (**f**), and the fractional shortening of the heart (**g**) in 12-month-old untreated mice and 18-month-old ATTAC mice treated with vehicle or AP. **h**, Fat mass ( $n = 9$  mice per group). 18-month-old

ATTAC vehicle-treated mouse values are the same as indicated in Fig. 1. **i**, iWAT and eWAT depot weight ( $n = 4$  mice per group). 18-month-old ATTAC vehicle-treated mouse values are the same as indicated in Fig. 1. **j**, **k**, Kidney sclerosis (**j**) and blood urea levels (**k**) ( $n = 4$  mice per group). 18-month-old ATTAC vehicle-treated mouse values are the same as indicated in Fig. 4. **l**, Time to death after isoproterenol administration ( $n = 4$  mice per group). 18-month-old ATTAC vehicle-treated mouse values are the same as indicated in Fig. 5. Error bars indicate s.e.m. No statistically significant differences were observed using unpaired two-tailed *t*-tests.





**Extended Data Figure 10 | Effect of senescent cell clearance on wound healing and tissue fibrosis.** **a**, Closure of 3-mm punch biopsy wounds in 18-month-old ATTAC females after treatment with vehicle or AP for 6 months and if drug treatment was stopped 2 days before skin puncture or continued during wound closure ( $n = 6$  wounds for -AP; -AP and +AP; -AP and  $n = 10$  wounds for -AP; +AP and +AP; +AP). AP administration during the wound healing process significantly attenuates the rate of wound closure independently of whether senescent cell removal had occurred before wounding. **b**, Closure of 3-mm punch biopsy wounds in 4-month-old ATTAC females after treatment with vehicle or AP following

wounding ( $n = 10$  wounds per group). Similar to 18-month-old mice, AP administration during the wound healing process dramatically attenuated the rate of wound closure. **c**, Quantification of total GFP<sup>+</sup> cells isolated from 3-mm punch biopsy wounds of 4-month-old mice two days into the wound healing process treated with vehicle (black) or AP (red,  $n = 3$  mice per group). **d**, PTAH-stained tissues sections from 18-month-old ATTAC mice for detection of fibrosis. Scale bars, 100  $\mu$ m. Error bars indicate s.e.m. Mice receiving AP during the healing process in **a** and **b** are significantly different from those treated with vehicle from day 1.5 to day 9.5.  $*P < 0.05$  (unpaired two-tailed  $t$ -tests).

# Multistage coupling of independent laser–plasma accelerators

S. Steinke<sup>1</sup>, J. van Tilborg<sup>1</sup>, C. Benedetti<sup>1</sup>, C. G. R. Geddes<sup>1</sup>, C. B. Schroeder<sup>1</sup>, J. Daniels<sup>1,3</sup>, K. K. Swanson<sup>1,2</sup>, A. J. Gonsalves<sup>1</sup>, K. Nakamura<sup>1</sup>, N. H. Matlis<sup>1</sup>, B. H. Shaw<sup>1,2</sup>, E. Esarey<sup>1</sup> & W. P. Leemans<sup>1,2</sup>

Laser-plasma accelerators (LPAs) are capable of accelerating charged particles to very high energies in very compact structures<sup>1</sup>. In theory, therefore, they offer advantages over conventional, large-scale particle accelerators. However, the energy gain in a single-stage LPA can be limited by laser diffraction, dephasing, electron-beam loading and laser-energy depletion<sup>1</sup>. The problem of laser diffraction can be addressed by using laser-pulse guiding<sup>2</sup> and preformed plasma waveguides to maintain the required laser intensity over distances of many Rayleigh lengths<sup>3</sup>; dephasing can be mitigated by longitudinal tailoring of the plasma density<sup>4</sup>; and beam loading can be controlled by proper shaping of the electron beam<sup>5</sup>. To increase the beam energy further, it is necessary to tackle the problem of the depletion of laser energy, by sequencing the accelerator into stages, each powered by a separate laser pulse<sup>6</sup>. Here, we present results from an experiment that demonstrates such staging. Two LPA stages were coupled over a short distance (as is needed to preserve the average acceleration gradient) by a plasma mirror. Stable electron beams from a first LPA were focused to a twenty-micrometre radius—by a discharge capillary-based<sup>7</sup> active plasma lens<sup>8</sup>—into a second LPA, such that the beams interacted with the wakefield excited by a separate laser. Staged acceleration by the wakefield of the second stage is detected via an energy gain of 100 megaelectronvolts for a subset of the electron beam. Changing the arrival time of the electron beam with respect to the second-stage laser pulse allowed us to reconstruct the temporal wakefield structure and to determine the plasma density. Our results indicate that the fundamental limitation to energy gain presented by laser depletion can be overcome by using staged acceleration, suggesting a way of reaching the electron energies required for collider applications<sup>6,9</sup>.

The limitations of conventional particle-accelerator technology<sup>10</sup> are motivating the development of advanced particle-acceleration techniques, such as laser-plasma acceleration, which could see a broad range of applications—ranging from particle colliders that produce energies beyond  $10^{12}$  electronvolts (TeV)<sup>9</sup> to compact free-electron lasers and Thomson  $\gamma$ -ray sources. Within the past few years, tremendous progress in LPA development has been made. After the first demonstration of per-cent-level energy spread and small divergence in millimetre-scale plasmas in 2004 (refs 11–13), electron beams with energies of  $10^9$  eV (GeV) were obtained with 40-terawatt laser pulses<sup>14</sup>. Subsequently, electron beams with multi-GeV energies were reported with petawatt-class laser systems and plasmas of a few centimetres in length<sup>15–17</sup>. Controlling the injection of electrons into plasma waves enables the accelerator to be precisely tuned<sup>18–21</sup>.

The accelerating gradient,  $E_z$ , of a single-stage LPA scales with the plasma density,  $n_e$ , as  $E_z \propto n_e^{1/2}$ . The single-stage length,  $L_{\text{stage}}$ , is given by the laser-depletion length,  $L_{\text{deplete}}$ :  $L_{\text{stage}} \approx L_{\text{deplete}} \propto n_e^{-3/2}$ . Thus, the energy gain per stage scales as  $W_{\text{stage}} \propto \frac{1}{n_e}$ . With the help of particle-in-cell simulations, it has been shown<sup>22</sup> that, in order to reach an energy

of 1 TeV in a single stage, a plasma density of about  $10^{15} \text{ cm}^{-3}$  is required. This would result in an acceleration length of 1 km, a low acceleration gradient, 10 kJ of required laser pulse energy, and an electron bunch that is not suitable for collider applications<sup>9</sup>. However, staging using multiple petawatt laser systems would allow for the use of much higher plasma densities, and hence the generation of higher accelerating gradients; this would result in a reduction in the total LPA-based linear-accelerator length to a few hundred metres, as well as more favourable laser parameters and electron-bunch charges<sup>9</sup>. To obtain such a compact set-up, coupling distances of the order of the laser-depletion length, at the 1-metre scale, are assumed. Because the fluence restrictions of conventional laser optics require them to be positioned several metres away from the focal plane of the laser, plasma mirrors<sup>23</sup> have been proposed instead as the final steering optics<sup>24</sup>. Such a compact staging set-up is also important to photon sources (for example when using  $\gamma$ -rays to inspect materials)<sup>25</sup>, where it can be used to decelerate electrons after photon production to mitigate shielding needs.

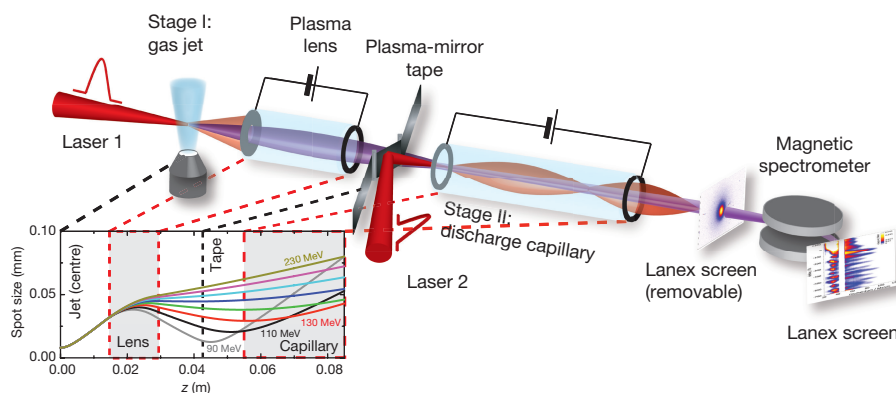
Here we demonstrate coupling of, and acceleration in, two separately powered LPA stages. Two synchronized laser pulses were applied to drive two acceleration stages in series (Fig. 1). The first stage generated electron beams from a gas-jet target with a central energy of 120 MeV (see Methods). To maximize the coupling efficiency to the second stage, these electron beams were refocused by a first discharge capillary, acting as an active plasma lens<sup>8</sup>, to the entrance of a second discharge capillary, serving as the second-stage target. The acceleration fields in the second capillary were excited by the second laser pulse, reflected by a tape-based plasma-mirror a few centimetres away. Depending upon the relative timing of the two laser pulses, an energy gain of about 100 MeV might be observed, with a charge-coupling efficiency of 3.5%. Continuous scanning of the relative timing of the laser pulses allowed us to reconstruct the femtosecond-scale temporal field structure of the second-stage wake, providing an important wake diagnostic. Numerical modelling confirms the effective trapping of the electron beam in the second-stage wake structure, and provides evidence for the femtosecond duration of the first-stage electron beams.

The electron beams generated in the first stage were transported to the second-stage target using a pulsed active plasma lens<sup>8</sup>. Radially symmetrical focusing was achieved in a gas-filled, 15-mm-long capillary with a diameter of 500  $\mu\text{m}$ , using an axial discharge current of 650 A, which produced an azimuthal focusing magnetic field. The high field strengths produced ( $\sim 0.5$  tesla) re-focused electrons with energies of 75–125 MeV within a distance of 25 mm through the plasma-mirror tape to an energy-dependent spot size of 20–30  $\mu\text{m}$  (r.m.s.) at the second plasma stage (Fig. 1 inset). The divergence acceptance of the lens was 5 mrad.

The second-stage LPA target was formed by a separate discharge capillary structure (see Methods). The discharge current created a pre-formed plasma that served as a waveguide, guiding the driving

<sup>1</sup>Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA. <sup>2</sup>University of California–Berkeley, Berkeley, California 94720, USA. <sup>3</sup>Eindhoven University of Technology, PO Box 513, 5600MB Eindhoven, The Netherlands.





**Figure 1 | The experimental set-up.** In stage I, a pulse of laser light is focused on a gas jet, producing an electron beam. This beam is then transported to the entrance of stage II by a discharge capillary, which is acting as an active plasma lens. In stage II, the beam enters a second discharge capillary. A second laser pulse further accelerates the electrons; this laser is coupled to the second discharge capillary via a plasma-mirror tape. Lanex screens are used to detect the energy integrated and

energy-dispersed (as part of a dipole spectrometer) electron profiles. The inset shows how the diameter of the waist (the ‘spot size’) of the electron beam evolves along the beam path ( $z$ ), simulated for different electron-beam energies produced by the first stage, according to ref. 8. Energies in the interval 75–125 MeV are focused at the entrance of the stage II capillary to spot sizes of the order of the input-laser spot size ( $18\ \mu\text{m}$ ).

laser pulse over many Rayleigh lengths, minimizing diffraction and extending the acceleration length. These target systems are well characterized<sup>3,14</sup>, and a model has previously been developed that permits the wakefield amplitude to be determined by means of the spectral redshift of the transmitted laser<sup>15,26,27</sup>. We used a feedback-controlled, tape-based plasma mirror (see Methods) to combine the injected electron beam with the laser driver in the second stage.

The laser pulses reflected off the plasma mirror were guided in the parabolic plasma channel created in the discharge capillary with an energy transmission of 85%. Matched propagation of a transversely Gaussian laser pulse in a plasma with a transverse parabolic density profile can be obtained—at low laser power and intensity—if the input-laser spot size,  $w_0$  (which corresponds to a radius whereby the laser intensity is  $1/e^2$  compared to the on-axis value), equals the matched spot size,  $r_m$ . (For a parabolic plasma profile,  $n(r) = n_0 + \alpha r^2$ , where  $n_0$  is the on-axis density,  $r$  is the transverse spatial coordinate in the plasma channel, and  $\alpha$  is the parameter controlling the depth of the channel; the matched spot size is given by  $r_m = (\alpha \pi r_e)^{-1/4}$ , with  $r_e = 2.8 \times 10^{-13}\ \text{cm}$  being the classical electron radius.) In our experimental conditions,  $r_m = 45\ \mu\text{m}$ , and the laser spot size at focus was  $w_0 = 18\ \mu\text{m}$ , leading to mismatched propagation and, hence, to varying peak intensities and wakefield strengths along the capillary. The characteristic oscillation length of the laser spot size is given by  $\lambda_{OS} = \pi z_{RM}$ , where  $z_{RM} = \pi r_m^2 / \lambda$ , and  $\lambda = 0.8\ \mu\text{m}$  is the central wavelength of the laser. For our parameters,  $\lambda_{OS} = 25\ \text{mm}$ . Wake excitation under these conditions was confirmed by measuring optical spectra of the transmitted laser pulse, showing an increasing redshift with increasing plasma density in the channel. Quantitative analysis of the spectra revealed a maximum relative redshift of 3% with respect to the central wavelength of the laser at a density of  $2 \times 10^{18}\ \text{cm}^{-3}$ . This corresponds to an average field amplitude of about  $17\ \text{MV mm}^{-1}$  if wake excitation occurs over the full length of the capillary<sup>26</sup>.

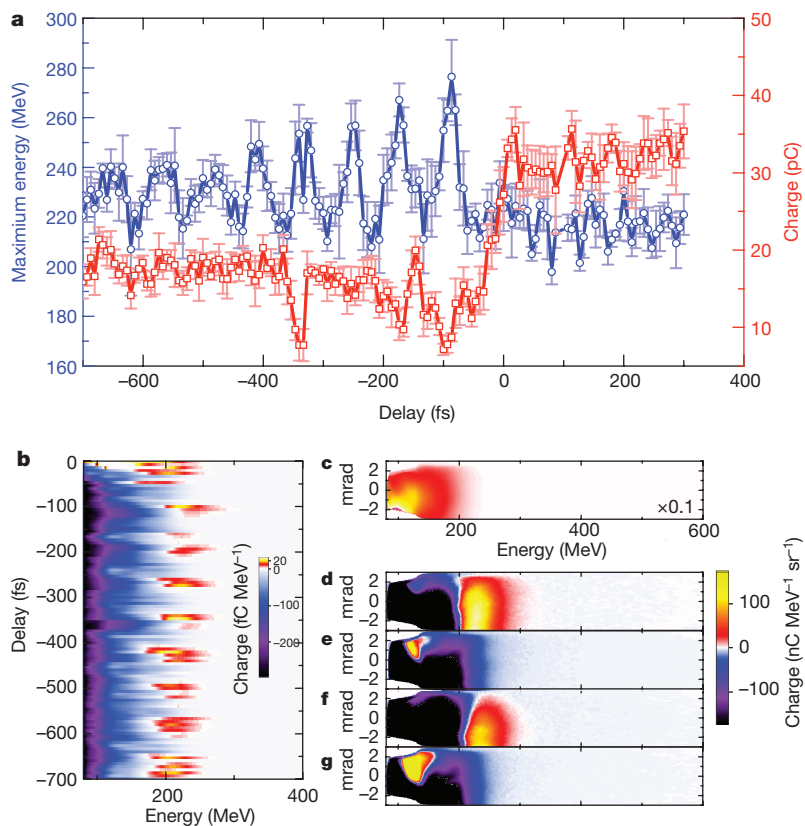
To control the phasing of the electron beam in the plasma wake of the second-stage LPA, we varied the delay between the laser pulses that drive the first and second stages, with femtosecond precision, with an optical-delay stage in the laser beam line of the injector stage. Electron spectra were recorded as a function of the delay between the two laser pulses. In the case of a positive delay, the first-stage electrons propagated without the influence of the second laser pulse. After the second laser pulse arrived (negative delay), the electron spectra were periodically modulated in energy (Fig. 2a). The period of the modulation was  $80 \pm 6$  femtoseconds, consistent with a plasma wavelength  $\lambda_p$  of  $24\ \mu\text{m}$ , at a density of  $(1.9 \pm 0.3) \times 10^{18}\ \text{cm}^{-3}$ . The constant periodicity of the observed modulation as a function of delay behind the driver pulse further indicates a quasilinear wake, consistent with

expectations for the experimental parameters, including laser intensity and plasma density.

To investigate the influence of the second-stage wakefield on the electron beam in detail, we subtracted the reference spectrum resulting from an unperturbed beam (positive delay) from the spectrum at each delay, to emphasize the effect of the second laser pulse while maintaining absolute charge information. The resulting electron distributions are plotted in Fig. 2b in the form of a waterfall plot of electron spectra, where each horizontal line corresponds to an energy spectrum that is averaged over five shots. Background-subtracted two-dimensional charge maps for the first two peaks and valleys of the blue curve in Fig. 2a, also averaged over five shots, are shown in Fig. 2d–g. The presence of the second-stage laser results in a reduction in total beam charge by up to a factor of three (Fig. 2a). For appropriate timing of the second-stage laser, however, charge was detected beyond the energy cut-off of the input electron spectrum, that is,  $>200\ \text{MeV}$ . This charge accelerated beyond the cut-off of the input spectrum (red and yellow areas in Fig. 2b, d, f), which indicates acceleration in the second stage. The integrated charge of  $1.2\ \text{pC}$  in this region represents the charge trapped in the accelerating phase of the wake, corresponding to a trapping efficiency of 3.5%. At delays of  $\lambda_p/2$  after the times of maximum energy gain, roughly  $1\ \text{pC}$  of additional charge was detected around  $110\text{--}150\ \text{MeV}$  (Fig. 2e, g). This could correspond to electrons that have decelerated, or to electrons that have been deflected by the transverse wake fields into the spectrometer acceptance. The broad energy spread of the first-stage electron beam prevents unambiguous observation of the decelerating phase of the wake under these conditions.

Numerical modelling performed with the code INF&RNO<sup>28,29</sup> allows detailed analysis of the interaction. Figure 3a shows reference-subtracted electron spectra as a function of the delay between the arrival of the electron bunch and the laser pulse. The simulations show that the observed energy modulations depend on the phasing of the electron bunch within the wake. The periodicity of the modulation is determined by the plasma density and is consistent with the experimental observation. However, the amount of post-accelerated charge decreases in the later accelerating phases of the wake as a result of increasing wake curvature. The fact that the linearity of the wake appears to be preserved in the experimental results could be attributed to a deviation from the parabolic plasma channel. We have found that, for example, simulating a quartic plasma density profile yields a charge distribution similar to that obtained in the experiment (Extended Data Fig. 1). Simulations performed assuming matched guiding conditions, and a more-energetic injector beam with reduced energy spread, indicate that roughly 90% trapping can be achieved (Extended Data Fig. 2).

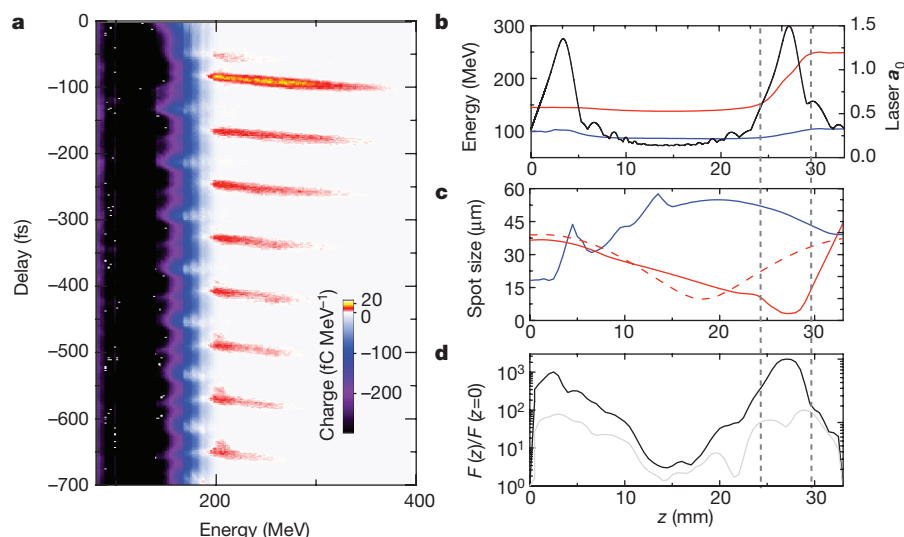




**Figure 2 | Spectra of electron beams produced by staged acceleration.** **a**, Maximum electron energy (blue) and total electron-beam charge (red) as a function of the delay between the two driving laser pulses. ‘Positive delays’ correspond to times before the arrival of laser 2. A single data point represents an average of five measurements; error bars represent the standard deviation. **b**, Waterfall plot of electron spectra (five-shot average), each with the reference from panel **c** subtracted, as a function of delay. **c**, 100-shot average unperturbed reference for delays of 100–300 fs before the arrival of the second laser pulse. **c–g**, Two-dimensional charge maps (five-shot average), with reference (**c**) subtracted for the first two maxima and minima of the energy oscillation shown in **a**—that is, for delays of –107 fs (**d**), –153 fs (**e**), –193 fs (**f**) and –240 fs (**g**). The y-axis in **c–g** shows the transverse angle in milliradians; ‘0’ corresponds to the laser axis.

As discussed above, there were two regions of increased laser intensity—and hence higher wake amplitude—in the capillary, owing to mismatched laser-pulse guiding. In Fig. 3, we plot the evolution of electron energy (Fig. 3b) and bunch size (Fig. 3c) as a function of propagation in the capillary for two different electron populations at a delay of –252 femtoseconds. Electrons that had an initial energy in the range 75–125 MeV did not gain a notable amount of energy, and

were strongly defocused by the transverse wakefield in the early stages of the laser–plasma interaction. On the other hand, electrons with a final energy above 200 MeV experienced an energy gain of about 100 MeV in the vicinity of the second laser focus, corresponding to a propagation distance of  $z = 24$ –29 mm in the plasma, where—because of the focusing induced by the discharge current and the laser-induced wake—they reach a spot size of roughly  $5\ \mu\text{m}$  and interact strongly



**Figure 3 | Simulation results.** **a**, Waterfall plot of electron energy spectra as a function of delay, with the same colour scale as in Fig. 2b. The reference was subtracted from each spectrum in a similar way as for the experimental results. **b**, Black line, evolution of the laser intensity, expressed as the relativistically normalized laser vector potential  $a_0$ , as a function of propagation in the capillary ( $z$  denotes the location along the plasma channel) in the second-stage LPA. Red and blue lines, evolution of electron energy as a function of propagation in the capillary for a delay of –252 fs for two different electron populations: electrons with an initial energy in the interval 75–125 MeV (blue) and electrons with a final energy

of 200–300 MeV (red). **c**, Evolution of the electron-beam r.m.s. spot size (also called bunch size) for the same electron-beam subsets as in panel **b**, along with the electron-beam r.m.s. spot size of the ‘red’ electrons without influence of the wakefield (red dashed line). **d**, Total transverse force on the electron beam divided by the contribution from the discharge current  $F(z)/F(z=0)$ , for two different distances from the axis:  $10\ \mu\text{m}$  (black) and  $20\ \mu\text{m}$  (grey). ( $F(z=0)$  is the contribution from the discharge current.) The area between the grey dashed lines in **b–d** indicates where the electrons with final energies greater than 200 MeV are trapped and accelerated in the laser-induced wakefield.

with the laser-driven plasma wake. This is shown in Fig. 3c, where we plot (red solid line) the evolution of the transverse electron bunch size as a function of the propagation distance. The red dashed line in Fig. 3c shows the evolution of the bunch spot size without the influence of the laser pulse, indicating that the contribution of the external magnetic field (induced by the discharge current of the second capillary) to the trapping of the electrons cannot be neglected.

To elucidate this, we computed in the simulation the amplitude of the transverse force,  $F(z)$ , acting on the electron bunch at distances of 10  $\mu\text{m}$  and 20  $\mu\text{m}$  from the capillary axis. The focusing contribution from the externally applied discharge current is given by  $F(z=0)$ , where there is no wake contribution. We plot the ratio  $F(z)/F(z=0)$  in Fig. 3d. For propagation distances in the capillary of  $z < 10$  mm, the average spot size of electrons with final energies of 200–300 MeV is larger than the laser spot; hence, the external field prevails and focuses the electron beam. At positions  $10 \text{ mm} < z < 22$  mm, the laser intensity is reduced markedly because of mismatched guiding (Fig. 3b); hence the focusing force from the externally applied current is of the same order as the wakefield contribution, leading to continued focusing of the electron beam. Finally, for positions  $24 \text{ mm} < z < 29$  mm, the contribution of the wakefield dominates the interaction. In this range, a subset of the electron beam is trapped in the wakefield and gains energy. An energy gain of about 100 MeV is observed in the simulations. For comparison, the experiment was repeated with the reversed (that is, defocusing) current in the second capillary. In this case, the external focusing field that supports trapping in the region  $24 \text{ mm} < z < 29$  mm is absent. Trapping still occurs in the region of the first laser focus ( $2 \text{ mm} < z < 5$  mm), but electrons trapped here are subsequently dephasing/defocusing and thus do not experience a detectable energy gain at the end of the second capillary. The full data set is shown in Extended Data Fig. 3.

In summary, we have presented an experimental study of the coupling of two LPA stages that are independently driven by two synchronized laser pulses. We demonstrated electron-beam injection and capture into the second-stage LPA by means of approximately a 100-MeV energy gain that recurs at delays corresponding to multiples of  $\lambda_p$ . The observation of temporally well defined energy modulations further implies directly that the bunch length of the input electron beam is shorter than  $\lambda_p/4 \approx 5 \mu\text{m}$ . This represents a milestone in the development of laser-driven, plasma-based accelerators for particle colliders, and for any other LPA application that requires electron energies beyond the limits of single stages. Simulations of modules that provide multi-GeV energy boosts to high-quality electron beams by operating at lower densities (that is, with a larger transverse wake size) suggest that 100% trapping can be achieved. The demonstration of such collider-relevant stages—including synchronizing the timing of separate laser drivers—will be the focus of our future work.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 24 September; accepted 27 November 2015.**

**Published online 1 February 2016.**

- Esarey, E., Schroeder, C. B. & Leemans, W. P. Physics of laser-driven plasma-based electron accelerators. *Rev. Mod. Phys.* **81**, 1229–1285 (2009).
- Durfee, C. G. & Milchberg, H. M. Light pipe for high intensity laser pulses. *Phys. Rev. Lett.* **71**, 2409–2412 (1993).
- Spence, D. J. & Hooker, S. M. Investigation of a hydrogen plasma waveguide. *Phys. Rev. E* **63**, 015401 (2000).
- Rittershofer, W., Schroeder, C. B., Esarey, E., Gruner, F. J. & Leemans, W. P. Tapered plasma channels to phase-lock accelerating and focusing forces in laser-plasma accelerators. *Phys. Plasmas* **17**, 063104 (2010).
- Schroeder, C. B., Benedetti, C., Esarey, E. & Leemans, W. P. Beam loading in a laser-plasma accelerator using a near-hollow plasma channel. *Phys. Plasmas* **20**, 123115 (2013).

- Leemans, W. P. & Esarey, E. Laser-driven plasma-wave electron accelerators. *Phys. Today* **62**, 44 (2009).
- Gonsalves, A. J., Rowlands-Rees, T. P., Brooks, B. H. P., van der Mullen, J. J. A. M. & Hooker, S. M. Transverse interferometry of a hydrogen-filled capillary discharge waveguide. *Phys. Rev. Lett.* **98**, 025002 (2007).
- van Tilborg, J. *et al.* Active plasma lensing for relativistic laser-plasma-accelerated electron beams. *Phys. Rev. Lett.* **115**, 184802 (2015).
- Schroeder, C. B., Esarey, E., Geddes, C. G. R., Benedetti, C. & Leemans, W. P. Physics considerations for laser-plasma linear colliders. *Phys. Rev. Special Topics Accel. Beams* **13**, 101301 (2010).
- Ellis, J. & Wilson, I. New physics with the compact linear collider. *Nature* **409**, 431–435 (2001).
- Geddes, C. G. R. *et al.* High-quality electron beams from a laser wakefield accelerator using plasma-channel guiding. *Nature* **431**, 538–541 (2004).
- Mangles, S. P. D. *et al.* Relativistic electron acceleration by a laser of intensity in excess of  $10(20) \text{ W cm}^{-2}$ . *Phys. Scr.* **T107**, 121–124 (2004).
- Faure, J. *et al.* A laser-plasma accelerator producing monoenergetic electron beams. *Nature* **431**, 541–544 (2004).
- Leemans, W. P. *et al.* GeV electron beams from a centimetre-scale accelerator. *Nature Phys.* **2**, 696–699 (2006).
- Leemans, W. P. *et al.* Multi-GeV electron beams from capillary-discharge-guided subpetawatt laser pulses in the self-trapping regime. *Phys. Rev. Lett.* **113**, 245002 (2014).
- Wang, X. M. *et al.* Quasi-monoenergetic laser-plasma acceleration of electrons to 2 GeV. *Nature Commun.* **4**, 1988 (2013).
- Kim, H. T. *et al.* Enhancement of electron energy to the multi-GeV regime by a dual-stage laser-wakefield accelerator pumped by petawatt laser pulses. *Phys. Rev. Lett.* **111**, 165002 (2013).
- Gonsalves, A. J. *et al.* Tunable laser plasma accelerator based on longitudinal density tailoring. *Nature Phys.* **7**, 862–866 (2011).
- Faure, J. *et al.* Controlled injection and acceleration of electrons in plasma wakefields by colliding laser pulses. *Nature* **444**, 737–739 (2006).
- Liu, J. S. *et al.* All-optical cascaded laser wakefield accelerator using ionization-induced injection. *Phys. Rev. Lett.* **107**, 035001 (2011).
- Amiranoff, F. *et al.* Observation of laser wakefield acceleration of electrons. *Phys. Rev. Lett.* **81**, 995–998 (1998).
- Vay, J. L. *et al.* Modeling of 10 GeV–1 TeV laser-plasma accelerators using Lorentz boosted simulations. *Phys. Plasmas* **18**, 123103 (2011).
- Thaury, C. *et al.* Plasma mirrors for ultrahigh-intensity optics. *Nature Phys.* **3**, 424–429 (2007).
- Sokollik, T. *et al.* Tape-drive based plasma mirror. *AIP Conf. Proc.* **1299**, 233–237 (2010).
- Powers, N. D. *et al.* Quasi-monoenergetic and tunable X-rays from a laser-driven Compton light source. *Nature Photon.* **8**, 29–32 (2014).
- Shiraishi, S. *et al.* Laser red shifting based characterization of wake field excitation in a laser-plasma accelerator. *Phys. Plasmas* **20**, 063103 (2013).
- Andreev, N. E. *et al.* Analysis of laser wakefield dynamics in capillary tubes. *New J. Phys.* **12**, 045024 (2010).
- Benedetti, C., Schroeder, C. B., Esarey, E., Geddes, C. G. R. & Leemans, W. P. Efficient modeling of laser-plasma accelerators with INF&RNO. *AIP Conf. Proc.* **1299**, 250–255 (2010).
- Benedetti, C., Schroeder, C. B., Esarey, E. & Leemans, W. P. Efficient modeling of laser-plasma accelerators using the ponderomotive-based code INF&RNO. *Proc. 11th International Computational Accelerator Physics Conference (Joint Accelerator Conferences Website) THAA12* (2012).

**Acknowledgements** We thank S. Shiraishi and T. Sokollik for their contributions to the initial construction of the set-up and an early version of the experiment, as well as C. Toth, D. Syversrud, N. Ybarrolaza, M. Kirkpatrick, G. Mannino, T. Sipla, D. Evans, R. Duarte, D. Baum and D. Munson for their contributions. This work was supported by the US Department of Energy, Office of Science, Office of High Energy Physics, under contract no. DE-AC02-05CH11231; by the US Department of Energy National Nuclear Security Administration, Defense Nuclear Nonproliferation R&D (NA22); and by the National Science Foundation (NSF) under contracts 0917687, 0935197, and PHY-1415596. This research used computational resources (Edison, Hopper) of the National Energy Research Scientific Computing Center (NERSC), which is supported by the Office of Science of the US Department of Energy under contract no. DE-AC02-05CH11231.

**Author Contributions** All authors contributed extensively to the work in this paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.P.L. ([wpleemans@lbl.gov](mailto:wpleemans@lbl.gov)).

## METHODS

Laser pulses with a centre wavelength,  $\lambda$ , of  $0.8\ \mu\text{m}$ , produced by the TREX Ti:sapph laser at the BELLA Center, were split with a  $\text{MgF}_2$  beamsplitter of 5-mm thickness. For stage I (2), 70% (30%) of the 40-TW laser-pulse energy—that is, 1.3 J (0.45 J) of laser energy—was focused by a parabola of focal length 2 m to a beam waist,  $w_0$ , of  $18\ \mu\text{m}$ . This corresponds to a laser intensity of  $4 \times 10^{18}\ \text{W cm}^{-2}$  ( $1.4 \times 10^{18}\ \text{W cm}^{-2}$ ) at a pulse duration of 45 fs.

**Plasma mirror.** A feedback-controlled, VHS-tape-based plasma-mirror system was used to combine the injection beam with the laser driver for the second stage. The electron beam passed through the tape, while the laser beam was reflected at  $90^\circ$ . This plasma mirror<sup>24</sup> has a laser-energy throughput of 80% (that is, 480 mJ laser energy), results in no observable laser mode or pointing degradations, and has been fully characterized and improved to provide uninterrupted operation at a repetition rate of 1 Hz for hours of run time.

**Electron-beam diagnostic.** The energy spectrum of electron bunches emerging from the plasma was measured by a magnetic electron spectrometer. It used a 120 mm by 480 mm rectangular-pole electromagnet with the maximum field strength of 1.5 T (ref. 30). Deflected electrons went through scintillating screens (Lanex, Kodak) that were arranged to optimize the energy resolution<sup>31</sup>, and the resulting light was imaged by three synchronously triggered charge-coupled-device (CCD) cameras, enabling single-shot detection of electrons with energies in the range 0.055–0.25 GeV and 0.3–2.1 GeV for 1 T.

**Gas jet.** A supersonic deLaval nozzle type gas jet with a diameter of  $700\ \mu\text{m}$  was used as the first-stage target. At backing pressures of 150 psi (with the gas being a mixture of 99% helium and 1% nitrogen), the neutral density profile exhibited an approximately Gaussian shape, verified by interferometric measurements<sup>32</sup>. The plasma density at the interaction point with the laser pulse was  $5 \times 10^{18}\ \text{cm}^{-3}$ . **Stage I electron beam.** To establish a stable injector stage, 70% of the energy of the BELLA Center TREX laser pulse was focused onto a supersonic gas jet (see above). A composition of two gases (99% helium and 1% nitrogen) was used to increase the amount of trapped charge. In such a mixed gas, injection was achieved by ionizing deeply bound electrons from a gas with high atomic number (nitrogen) near the peak of the laser pulse, the proper phase inside the wakefield allowing them to be trapped<sup>33,34</sup>. Stable electron beams with a mean energy of  $120 \pm 5\ \text{MeV}$ , a 60% (full width at half-maximum) energy spread, a beam charge of  $33 \pm 5\ \text{pC}$ , and a low divergence of  $4 \pm 0.3\ \text{mrad}$  were routinely produced, with a pointing stability of 0.3 mrad (standard deviation) over hours of run time (corresponding to thousands of laser shots and more than ten days).

**Capillaries.** The capillary used as the active plasma lens had a diameter of  $500\ \mu\text{m}$  and a length of 15 mm. Hydrogen gas was flowed into the capillary through slots located 2 mm from each end of the capillary. The temporal profile of the discharge current was approximately sinusoidal, with a peak current of 650 A. Data were taken for laser arrival at the peak of the discharge, and allowed focusing of the 100-MeV electrons to the entrance of the accelerator-stage capillary, which was located 25 mm downstream of the exit of the active plasma lens. The capillary used as the second stage had diameter of  $250\ \mu\text{m}$  and length of 33 mm. Hydrogen gas was flowed into the capillary through slots located 2 mm from each end of the capillary. The discharge current was approximately sinusoidal, with the peak current being 325 A. Data were taken for laser arrival at the peak of the discharge.

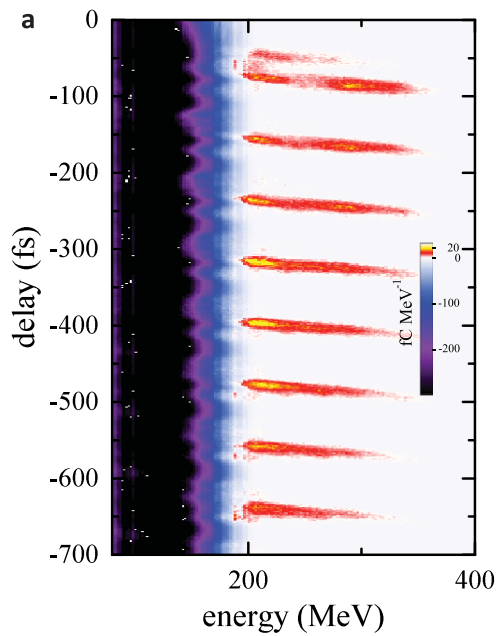
The two discharge capillaries have identical lens strength and form an imaging telescope for the electron beam. We verified independently that an electron propagating through this imaging telescope, without the influence of the second-stage laser pulse, has the same final divergence as the input electron beam, in the range of 60 to 300 MeV that is relevant for this experiment.

**Numerical modelling.** Calculations of the laser–plasma interaction—including laser and wakefield evolution—and of electron acceleration/modulation were carried out with the computational framework INF&RNO<sup>28</sup> (integrated fluid and particle simulation code). INF&RNO is a two-dimensional, cylindrical ( $r-z$ ) code that adopts an envelope model for the laser pulse, and makes use of the approximation of the ponderomotive force to describe the interaction of the laser pulse with the plasma. In the simulations presented here, the background plasma is modelled via a (noiseless) cold fluid description, while a fully kinetic description, realized through the particle-in-cell approach, is used for the electron bunch. The laser envelope is evolved in time using a second-order Crank–Nicolson method. The full wave operator is retained in the evolution equation for the laser envelope. At any given time step, once the laser envelope is specified, the (fluid) wakefield is computed using a quasistatic approach. The effect of the discharge current in stage II is taken into account by adding a radially varying azimuthal magnetic field corresponding to a discharge current of 325 A. Evolution of the electron bunch over time is computed using a fourth-order Runge–Kutta integrator. The input laser pulse is modelled according to the measured transverse intensity profile of the laser. The background plasma profile in stage II has a length of 33 mm, with 3-mm initial/final ramps and a 27-mm-long longitudinal flat-top section in the centre. The transverse density profile is parabolic with a matched radius,  $r_m$ , of  $45\ \mu\text{m}$ . The initial (that is, after the first LPA stage) phase space distribution for the electron bunch is modelled according to the measured bunch properties (see the spectrum in Fig. 2c). The initial bunch length was varied in the range  $1\ \mu\text{m}$  to  $30\ \mu\text{m}$  (longitudinal flat-top profile), while the initial bunch radius was chosen in the range  $0.2\text{--}5\ \mu\text{m}$  (transverse flat-top profile). Simulation results are insensitive to the choice of the initial bunch radius within the aforementioned range. In each simulation, the complete transport of the electron bunch from the exit of stage I to the entrance of stage II is modelled, including the effect of the active plasma lens (powered with a current of 650 A) and the interaction (scattering) with the plasma-mirror tape. The final (that is, after stage II) electron spectra have been computed taking into account the finite acceptance of the spectrometer.

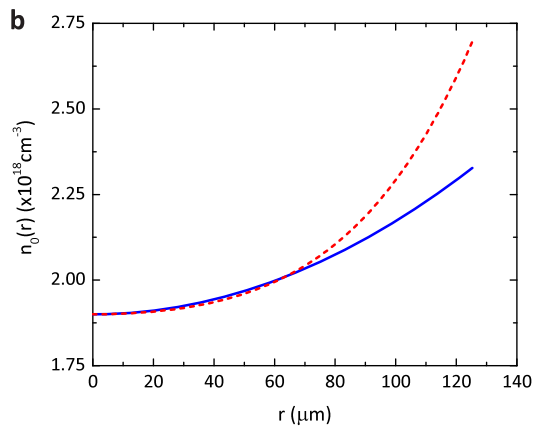
**Code availability.** The INF&RNO code is part of BLAST (Berkeley Lab Accelerator Simulation Toolkit). For more information on how to access the code, visit <http://blast.lbl.gov/>.

30. Nakamura, K. *et al.* Broadband single-shot electron spectrometer for GeV-class laser-plasma-based accelerators. *Rev. Sci. Instrum.* **79**, 053301 (2008).
31. Nakamura, K. *et al.* Electron beam charge diagnostics for laser plasma accelerators. *Phys. Rev. Special Topics Accel. Beams* **14**, 062801 (2011).
32. Plateau, G. R. *et al.* Wavefront-sensor-based electron density measurements for laser-plasma accelerators. *Rev. Sci. Instrum.* **81**, 033108 (2010).
33. Rowlands-Rees, T. P. *et al.* Laser-driven acceleration of electrons in a partially ionized plasma channel. *Phys. Rev. Lett.* **100**, 105005 (2008).
34. Chen, M., Esarey, E., Schroeder, C. B., Geddes, C. G. R. & Leemans, W. P. Theory of ionization-induced trapping in laser-plasma accelerators. *Phys. Plasmas* **19**, 033101 (2012).



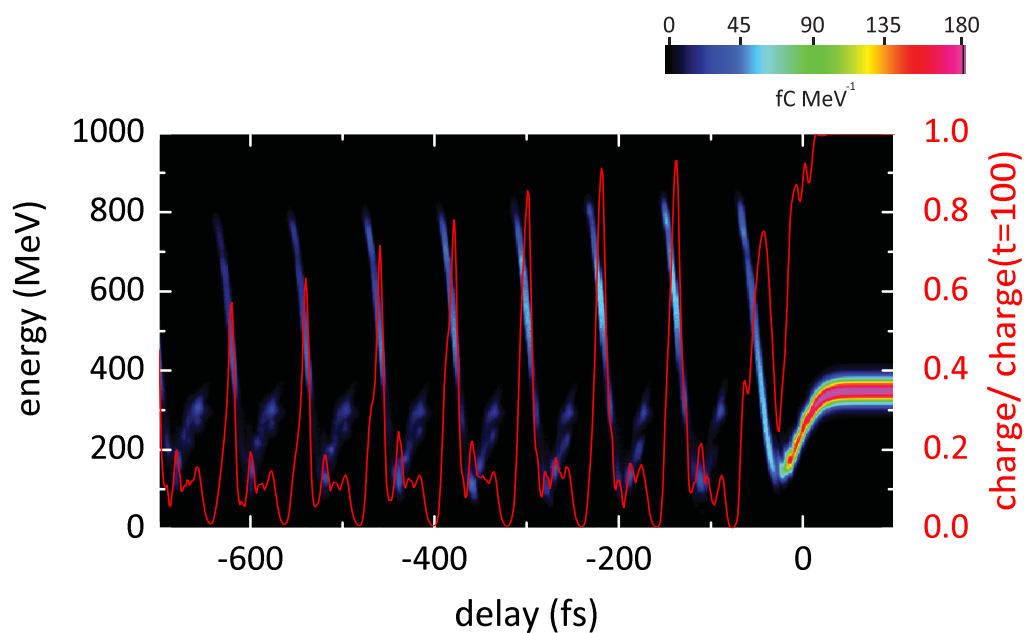


**Extended Data Figure 1 | Simulation results with quartic transverse plasma density profile.** **a**, Waterfall plot of electron energy spectra as a function of delay, with the same colour scale as in Fig. 2b. The reference spectrum (that is, the spectrum computed for delays greater than 0, without the influence of laser 2) was subtracted from each simulation spectrum in a similar way as for the experimental results in Fig. 2b. **b**, The quartic transverse density profile used in the simulation is



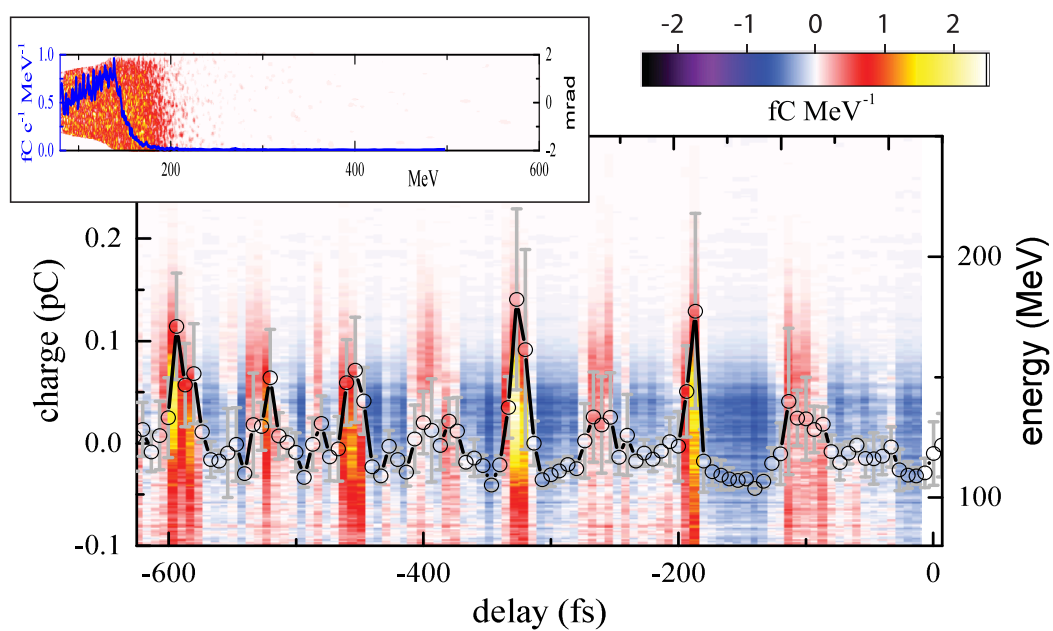
$$n(r) = n_0 + \varepsilon \alpha r^2 + \frac{\alpha}{2} (1 - \varepsilon) (\pi \alpha r_c)^{1/2} r^4, \text{ with } \varepsilon = 0.7 \text{ (red dashed line).}$$

The chosen form for the quartic profile is such that for any chosen  $\varepsilon$  ( $0 \leq \varepsilon \leq 1$ ), the matched spot size in the limit of a low laser power and low laser intensity is the same as for the parabolic profile  $\varepsilon = 1$  (blue line).  $n_0$  is the on-axis density,  $r$  is the transverse (radial) coordinate, and  $\alpha$  is the parameter controlling the depth of the plasma channel.



**Extended Data Figure 2 | Simulation with optimized laser–plasma parameters.** Waterfall plot of electron energy spectra as a function of delay. This simulation assumed matched laser-guiding conditions, and a more-energetic injector beam with reduced energy spread (as compared with the experimental set-up); the simulation results indicate that roughly

90% trapping of the electron beam can be achieved. The injector electron beam had a central energy of 350 MeV, a charge of 10 pC, an energy spread of 6% (r.m.s.) and a divergence of 2.5 mrad (full width at half-maximum). The laser pulse energy was 1 J. The laser spot size was  $w_0 = 40 \mu\text{m}$ . The red line shows the fraction of trapped charge (scale on the right).



**Extended Data Figure 3 | Experimental results obtained with a reversed (defocusing) current direction in the second LPA stage.** Main figure, a waterfall plot of electron spectra (subtracted by the reference) as a function of delay; the electron charge density is colour coded, and the total energy of the electron beam is shown as black circles. Inset, the

unperturbed reference beam. Electrons trapped in the region of the first laser focus will subsequently dephase/defocus under these conditions, and thus do not experience a detectable energy gain at the end of the second capillary.



# Observing the Rosensweig instability of a quantum ferrofluid

Holger Kadau<sup>1</sup>, Matthias Schmitt<sup>1</sup>, Matthias Wenzel<sup>1</sup>, Clarissa Wink<sup>1</sup>, Thomas Maier<sup>1</sup>, Igor Ferrier-Barbut<sup>1</sup> & Tilman Pfau<sup>1</sup>

Ferrofluids exhibit unusual hydrodynamic effects owing to the magnetic nature of their constituents. As magnetization increases, a classical ferrofluid undergoes a Rosensweig instability<sup>1</sup> and creates self-organized, ordered surface structures<sup>2</sup> or droplet crystals<sup>3</sup>. Quantum ferrofluids such as Bose–Einstein condensates with strong dipolar interactions also display superfluidity<sup>4</sup>. The field of dipolar quantum gases is motivated by the search for new phases of matter that break continuous symmetries<sup>5,6</sup>. The simultaneous breaking of continuous symmetries such as the phase invariance in a superfluid state and the translational symmetry in a crystal provides the basis for these new states of matter. However, interaction-induced crystallization in a superfluid has not yet been observed. Here we use *in situ* imaging to directly observe the spontaneous transition from an unstructured superfluid to an ordered arrangement of droplets in an atomic dysprosium Bose–Einstein condensate<sup>7</sup>. By using a Feshbach resonance to control the interparticle interactions, we induce a finite-wavelength instability<sup>8</sup> and observe discrete droplets in a triangular structure, the number of which grows as the number of atoms increases. We find that these structured states are surprisingly long-lived and observe hysteretic behaviour, which is typical for a crystallization process and in close analogy to the Rosensweig instability. Our system exhibits both superfluidity and, as we show here, spontaneous translational symmetry breaking. Although our observations do not probe superfluidity in the structured states, if the droplets establish a common phase via weak links, then our system is a very good candidate for a supersolid ground state<sup>9–11</sup>.

Research in condensed matter physics is driven by the discovery of novel phases of matter, in particular, phases simultaneously displaying different types of order. A prime example is the supersolid state, which features crystalline order and superfluidity simultaneously<sup>9–11</sup>. This state has been elusive and claims of its discovery in helium<sup>12,13</sup> have been withdrawn recently<sup>14</sup>. One of the requirements for spontaneous spatially ordered structure formation in the ground state of a many-body system is the existence of long-range interactions such as those present in ferrofluids. As a consequence of these interactions, a magnetized ferrofluid forms stationary surface waves due to competition between gravitational or magnetic trapping, dipolar interaction and surface tension. This effect is known as the normal-field instability or Rosensweig instability<sup>2</sup> and leads to stable droplet patterns on a superhydrophobic surface<sup>3</sup>. For ferrofluids, the dispersion relation of surface excitations has a minimum at finite momentum, resembling the well-studied roton spectrum in liquid helium<sup>15</sup>. However, the physical interpretation of this minimum is very different for ferrofluids and helium. The origin of the dispersion-relation minimum for a ferrofluid is an energy gain due to the attractive part of the dipolar interaction resulting in a clustering of polarized dipoles in a head-to-tail configuration in periodic structures. Such roton-induced structures have also been discussed for quantum ferrofluids<sup>8,16</sup>. In close similarity with a classical ferrofluid, a competition exists between the harmonic trapping, dipolar interaction and contact

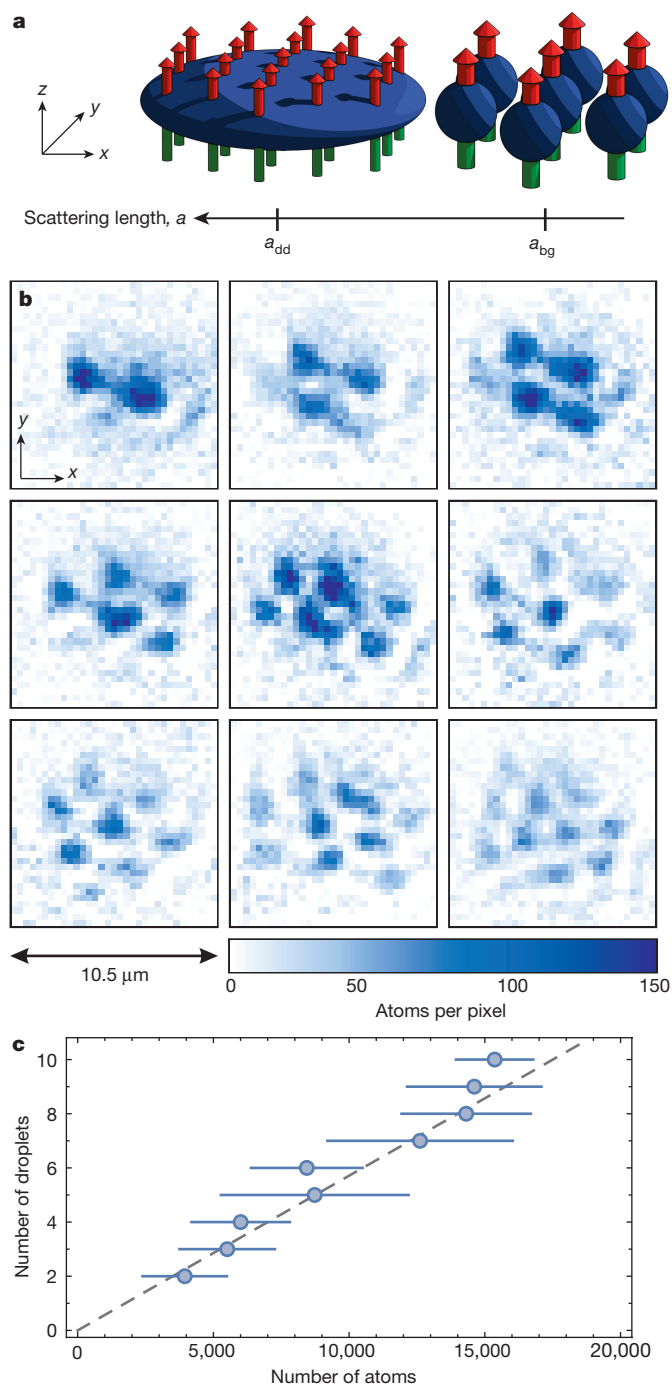
interaction in a quantum ferrofluid. For increasing relative dipolar interaction, the roton instability can lead to a periodic perturbation of the atomic density distribution, which is closely connected to the Rosensweig instability<sup>17</sup>. However, it was believed that these rotonic structures would be unstable, owing to subsequent instabilities of the forming droplets<sup>18</sup>.

Here we cool down the most magnetic element—dysprosium (Dy)<sup>19</sup>, with a magnetic moment of  $\mu = 9.93\mu_B$ , where  $\mu_B$  is the Bohr magneton—and generate a Bose–Einstein condensate (BEC)<sup>7</sup>. We observe an angular roton instability<sup>16,18</sup> and find subsequent droplet formation yielding triangular structures with surprisingly long lifetimes. We use two key tools to study these self-organized structures. First, we use a magnetic Feshbach resonance<sup>20</sup> to tune the contact interaction (see Extended Data Fig. 1) and to induce the droplet formation. Second, we use a microscope with high spatial resolution to detect the atomic density distribution *in situ*.

The first prediction of structured ground states in a dipolar BEC dates back to the early days of quantum gases<sup>21</sup>; the first mechanical effects were seen with chromium atoms<sup>22</sup>. There, the dipolar attraction deforms the compressible gas and its shape is balanced by a repulsive contact interaction, described by the scattering length  $a$ . To compare the strengths of the contact and dipolar interaction, we introduce a length scale that characterizes the magnetic dipole–dipole interaction strength:  $a_{dd} = \mu_0 \mu^2 m / (12\pi\hbar)$  (ref. 2), where  $\mu_0$  is the vacuum permeability,  $m$  is the atomic mass and  $\hbar$  is the reduced Planck constant. By tuning the scattering length  $a$  with a Feshbach resonance such that  $a < a_{dd}$ , the dipolar attraction dominates the repulsive contact interaction and an instability of a dipolar gas can occur<sup>4,23</sup>. However, in a pancake-shaped trap, the dipoles sit mainly side-by-side and predominantly repel each other, owing to the anisotropy of the dipole–dipole interaction, and hence the dipolar BEC is stabilized. In such a pancake-shaped configuration, the roton instability, which occurs at a finite wavelength, is predicted<sup>8,16</sup>.

For our experiments, we used the isotope <sup>164</sup>Dy with a dipolar length of  $a_{dd} = 132a_0$ , where  $a_0$  is the Bohr radius. This dipolar length is already greater than the background scattering length  $a_{bg} = 92(8)a_0$  (where the errors in parentheses here and elsewhere represent one standard deviation), which is the value far from Feshbach resonances<sup>24,25</sup>. To obtain a stable BEC, we tuned the scattering length to  $a \approx a_{dd}$  using a magnetic field of  $B = 6.962(3)$  G in the vicinity of a Feshbach resonance located at  $B_0 = 7.117(3)$  G. We then obtained typically 15,000 atoms in nearly pure Dy BECs (see Methods section). The atoms were trapped in a radially symmetric, pancake-shaped trap with harmonic frequencies of  $(\nu_x, \nu_y, \nu_z) = (46, 44, 133)$  Hz, and the external magnetic field aligned the magnetic dipoles in the axial  $z$  direction. Subsequently, we tuned the magnetic field such that  $B \lesssim 6.9$  G, which reduced  $a$  to  $a_{bg} < a_{dd}$ , resulting in an angular roton instability<sup>16</sup> that triggered the transition to ordered states (Fig. 1a). We then observed the formation of droplets that arranged in ordered structures using *in situ* phase-contrast polarization imaging along the  $z$  direction with a spatial resolution of  $1\mu\text{m}$ .

<sup>1</sup>Physikalisches Institut and Center for Integrated Quantum Science and Technology, Universität Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany.



**Figure 1 | Growth of a microscopic droplet crystal.** **a**, Schematic of the experimental procedure. We prepared a stable, strongly dipolar Dy BEC with  $a \approx a_{dd}$  in a pancake-shaped trap (left). By decreasing the scattering length  $a$ , we induced an instability close to  $a \approx a_{bg}$ . Following this instability, the atoms clustered to droplets in a triangular pattern (right). **b**, Representative single samples of droplet patterns imaged *in situ*, with droplet numbers,  $N_d$ , ranging from two to ten. **c**, We used a set of 112 realizations with different numbers of droplets and atoms for a statistical analysis. The plot shows the mean number of atoms as a function of the number of droplets  $N_d$ , with error bars indicating the standard deviation. The fitted linear relation (grey dashed line) has a slope of 1,750(300) atoms per droplet. This shows that increasing the number of atoms results in growth of the microscopic droplet crystal.

In Fig. 1b, we show typical *in situ* images of the resultant triangular patterns for the quantum ferrofluid with different numbers of droplets,  $N_d$ , ranging from two to ten. To analyse the average number of atoms

per droplet, we count the number of droplets  $N_d$  in relation to the total number of atoms. Figure 1c indicates a linear dependence between  $N_d$  and the number of atoms, with a slope of 1,750(300) atoms per droplet. For  $N_d = 2$ , we observe a droplet distance of  $d = 3.0(4) \mu\text{m}$ . The droplets, which have a large effective dipole moment of  $N_d \mu$ , strongly repel each other while the radial trapping applies a restoring force. Hence, the distance  $d$  can be calculated using a simplified one-dimensional classical system by minimizing the energy of the system.

We assume two strongly dipolar particles with 1,750 times the mass and magnetic moment of a Dy atom that are confined in a harmonic trap. For our experimental parameters, these particles minimize their energy with a distance of  $d = 3.3 \mu\text{m}$ , in agreement with the observed distance. For  $N_d > 2$ , the droplets arrange mostly in triangular structures, and form a microscopic crystal with a droplet distance of  $d = 2\text{--}3 \mu\text{m}$ . Owing to the isotropy of the repulsion between droplets in the radial plane, we expect the triangular configuration to have the lowest energy. Because of the repelling dipolar force between the droplets, we observe in the radial direction nearly round, discrete droplets with possible weak overlap to neighbouring ones.

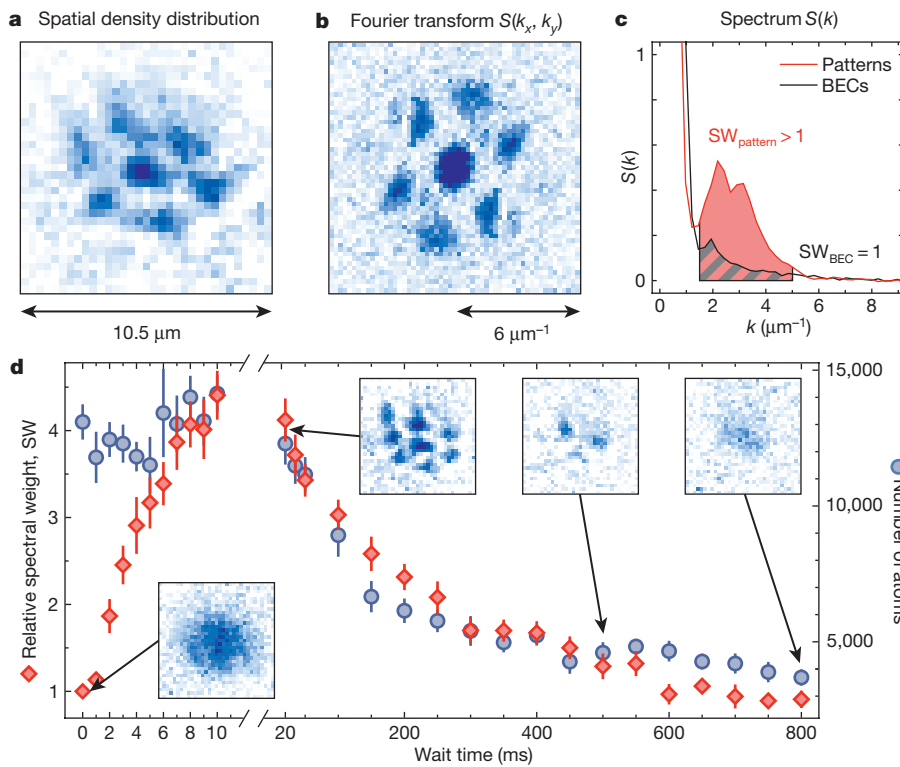
Comparing our quantum ferrofluid with a classical ferrofluid, very similar behaviour and patterns have been observed on a superhydrophobic surface<sup>3</sup>. In this classical-ferrofluid system, a single droplet first deforms as the external magnetic field increases, and then divides into two droplets when some critical field is reached. For a quantum ferrofluid, a single droplet should be unstable for  $a < a_{dd}$ , owing to the attractive part of the dipolar interaction, and so should collapse. Although, the counteracting quantum pressure—the zero-point energy that exists as a result of an external trapping potential—can compensate attraction and prevent collapse<sup>26</sup>, mean field calculations<sup>18</sup> predict this not to be the case. Our observation of stable droplet ensembles is therefore striking, and further work is needed to understand their stability. A possible stabilizing effect is that of quantum fluctuations, leading to beyond mean-field effects<sup>27</sup>. Such stabilization has been suggested in a similar situation of competing attraction and repulsion<sup>28</sup>, and an increased effect of quantum fluctuations has been calculated for strongly dipolar gases<sup>29</sup>.

As further quantitative statistical analysis, we computed the Fourier spectrum  $S(k)$  of the obtained images (Fig. 2a–c). The patterns are visible as a local maximum in  $S(k)$  at finite momentum  $k = 2\pi/d \approx 2.5 \mu\text{m}^{-1}$ , whereas the spectrum of a BEC monotonically decreases with  $k$ . We define the spectral weight

$$\text{SW} = \sum_{k=1.5 \mu\text{m}^{-1}}^{5 \mu\text{m}^{-1}} S(k)$$

which is a quantity that represents the strength of the structured states, and normalize it such that a BEC has  $\text{SW}_{\text{BEC}} = 1$ . After a quench of the interactions from  $a \approx a_{dd}$  to  $a \approx a_{bg}$ , we statistically investigated the pattern-formation time and the lifetime of these patterns (Fig. 2d). We repeated this measurement 13 times and found statistically that the pattern is fully developed after 7 ms, and has a 1/e lifetime of about 300 ms. The decay of the droplet structure is accompanied by a decrease in the number of atoms, with a 1/e lifetime of about 130 ms, while the residual thermal cloud is constant. Owing to the decreasing number of atoms, the structures evolve back to lower numbers of droplets,  $N_d$ , until they merge back into one droplet (insets of Fig. 2d). In comparison, because we measured lifetimes of a non-structured BEC of more than 5 s, we assume increased three-body losses as a reason for the reduced lifetime. One indication of this is the measured atomic peak density for droplets of  $n \gtrsim 5 \times 10^{20} \text{m}^{-3}$ , which is greater than the density of a BEC,  $n \approx 10^{20} \text{m}^{-3}$ .

To explore the nature of this instability, we performed the following experimental sequence, depicted in Fig. 3a. We prepared the BEC close to the Feshbach resonance with  $a \approx a_{dd}$  and ramped the magnetic field linearly to varying values near the instability point. We ensured that the structures were formed within 10 ms, even for values of the magnetic

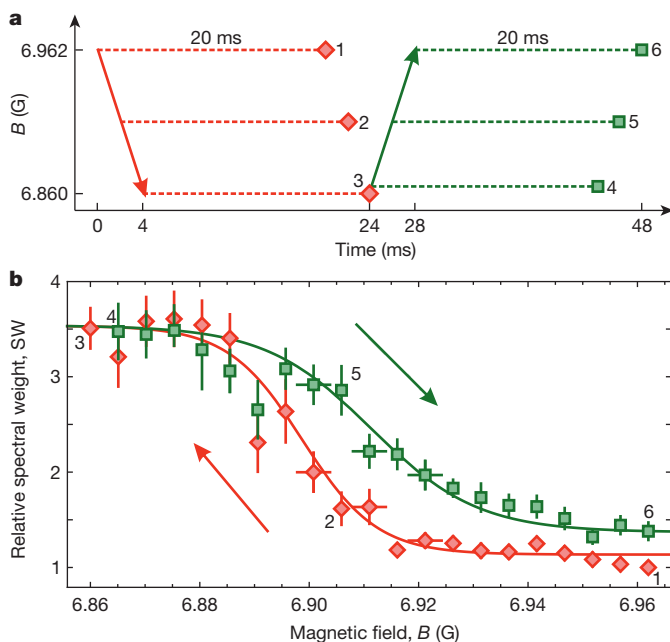


**Figure 2 | Evaluation of the structures and lifetime analysis.** **a–c**, Illustration of our statistical evaluation procedure. We began with the spatial density distribution (**a**), and then calculated the absolute value of the two-dimensional Fourier transform  $S(k_x, k_y)$  (**b**) and radially averaged over  $k = (k_x^2 + k_y^2)^{1/2}$  to obtain the spectrum  $S(k)$  (**c**). We removed the white noise from the spectrum  $S(k)$  such that  $S(k) = 0$  for  $k > 7 \mu\text{m}^{-1}$ , which corresponds to structures below our resolution. The spectra in **c** represent an average of 13 images for both BECs and patterns at a wait time of 10 ms. For patterns in the spatial density distribution, we observe an enhanced signal for  $k \approx 2.5 \mu\text{m}^{-1}$  in the spectrum (red line), whereas the spectrum of BECs (black line) shows monotonic decay for increasing momentum. We define the sum of these spectra over a momentum range as the relative spectral weight SW (shaded areas, as defined in the text), which is a quantity for the strength of the structured states. **d**, We performed a sudden quench (over 0.5 ms) of the BEC below the instability to  $B = 6.656(3)$  G, for varying wait times. To determine the creation time and the lifetime of the patterns, we plot the relative spectral weight SW (red diamonds) against wait time. Each point is an average of 13 realizations, with error bars indicating the standard error. The plot shows rapid pattern formation within 7 ms and a surprisingly long  $1/e$  lifetime of about 300 ms. This lifetime seems to be limited by a decrease in the number of atoms (blue circles). The insets are typical spatial density distributions of single samples before pattern formation (BEC; bottom left), and at three different wait times.

field close to the stability threshold, and waited at the chosen value for twice this time. Figure 3b shows a clear hysteresis for the way down in magnetic field compared to the return. For the return, we observe the same spectral weights as for the way down, but for magnetic fields that are about 20 mG greater. This demonstrates that our system features bistability in the transition region. We expect that the transition from one state to the other is driven by thermal excitations or weak currents due to overlapping droplets. In the thermodynamic limit, such

behaviour is a clear signature of a first-order phase transition and a latent heat in the crystallization process.

To verify that we are dealing with a transition to a ground state and not a metastable state resulting from quench dynamics, we performed forced evaporative cooling at a constant magnetic field far away from any Feshbach resonance with  $a \approx a_{bg}$ . We observed very similar self-organized structures, which were visible for temperatures near the expected critical temperature for the phase transition to a BEC.



**Figure 3 | Hysteresis of pattern formation.** **a**, A timeline of the experiment in which we observed hysteresis. We prepared the Dy BEC close to the Feshbach resonance at  $B = 6.962(3)$  G and ramped down the magnetic field linearly to different values, with a constant change rate, the lowest of which was  $B = 6.860(3)$  G. To ensure that the structures had enough time to form, we waited for 20 ms at each value before imaging the atomic sample *in situ*. For the way back, we first waited at the lowest field value for 20 ms and then increased the magnetic field with the same ramp speed to higher values, once again holding for 20 ms at each value before imaging the sample *in situ*. **b**, Hysteresis plot for the structured patterns, which shows the spectral weight SW against magnetic field as it is decreased (red diamonds and line) and then increased (green squares and line). Each point is an average of 14 realizations, with the vertical error bars indicating standard errors. We determined the long-term field stability to be 3 mG, as shown by the horizontal error bars for selected points. A clear hysteresis is visible, although the total time is twice as long for the way back. The labels 1–6 in **a** and **b** indicate data points at particular field values to help the understanding; the lines in **b** serve as a guide to the eye.



Because structures can melt back into a BEC, and because we observed evaporative cooling to patterns, it is quite plausible that the droplets are superfluid individually. Whether they share the same phase via weak links or lose their mutual phase coherence will have to be investigated in future experiments. Another open question is the creation dynamics of a self-organized structure. It will be interesting to study phonons in such a droplet crystal, which we expect to have eigenfrequencies of the order of many inverse lifetimes. We also expect that these phonons will be coupled to collective Josephson oscillations via weak links<sup>30</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 12 August; accepted 17 November 2015.**

**Published online 1 February 2016.**

- Rosensweig, R. *Ferrohydrodynamics* Ch. 7.1 (Cambridge Univ. Press, 1985).
- Cowley, M. D. & Rosensweig, R. E. The interfacial stability of a ferromagnetic fluid. *J. Fluid Mech.* **30**, 671–688 (1967).
- Timonen, J. V. I., Latikka, M., Leibler, L., Ras, R. H. A. & Ikkala, O. Switchable static and dynamic self-assembly of magnetic droplets on superhydrophobic surfaces. *Science* **341**, 253–257 (2013).
- Lahaye, T. *et al.* Strong dipolar effects in a quantum ferrofluid. *Nature* **448**, 672–675 (2007).
- Lahaye, T., Menotti, C., Santos, L., Lewenstein, M. & Pfau, T. The physics of dipolar bosonic quantum gases. *Rep. Prog. Phys.* **72**, 126401 (2009).
- Baranov, M. A., Dalmonte, M., Pupillo, G. & Zoller, P. Condensed matter theory of dipolar quantum gases. *Chem. Rev.* **112**, 5012–5061 (2012).
- Lu, M., Burdick, N. Q., Youn, S. H. & Lev, B. L. Strongly dipolar Bose-Einstein condensate of dysprosium. *Phys. Rev. Lett.* **107**, 190401 (2011).
- Santos, L., Shlyapnikov, G. V. & Lewenstein, M. Roton-Maxon spectrum and stability of trapped dipolar Bose-Einstein condensates. *Phys. Rev. Lett.* **90**, 250403 (2003).
- Andreev, A. F. & Lifshitz, I. M. Quantum theory of defects in crystals. *Sov. Phys. JETP* **29**, 1107–1113 (1969).
- Chester, G. V. Speculations on Bose-Einstein condensation and quantum crystals. *Phys. Rev. A* **2**, 256–258 (1970).
- Leggett, A. J. Can a solid be “superfluid”? *Phys. Rev. Lett.* **25**, 1543–1546 (1970).
- Kim, E. & Chan, M. H. W. Probable observation of a supersolid helium phase. *Nature* **427**, 225–227 (2004).
- Kim, E. & Chan, M. H. W. Observation of superflow in solid helium. *Science* **305**, 1941–1944 (2004).
- Kim, D. Y. & Chan, M. H. W. Absence of supersolidity in solid helium in porous Vycor glass. *Phys. Rev. Lett.* **109**, 155301 (2012).
- Henshaw, D. G. & Woods, A. D. B. Modes of atomic motions in liquid helium by inelastic scattering of neutrons. *Phys. Rev.* **121**, 1266–1274 (1961).
- Ronen, S., Bortolotti, D. C. E. & Bohn, J. L. Radial and angular rotons in trapped dipolar gases. *Phys. Rev. Lett.* **98**, 030406 (2007).
- Saito, H., Kawaguchi, Y. & Ueda, M. Ferrofluidity in a two-component dipolar Bose-Einstein condensate. *Phys. Rev. Lett.* **102**, 230403 (2009).
- Wilson, R. M., Ronen, S. & Bohn, J. L. Angular collapse of dipolar Bose-Einstein condensates. *Phys. Rev. A* **80**, 023614 (2009).
- Maier, T., Kadau, H., Schmitt, M., Griesmaier, A. & Pfau, T. Narrow-line magneto-optical trap for dysprosium atoms. *Opt. Lett.* **39**, 3138–3141 (2014).
- Maier, T. *et al.* Emergence of chaotic scattering in ultracold Er and Dy. *Phys. Rev. X* **5**, 041029 (2015).
- Góral, K., Rzażewski, K. & Pfau, T. Bose-Einstein condensation with magnetic dipole-dipole forces. *Phys. Rev. A* **61**, 051601 (2000).
- Stuhler, J. *et al.* Observation of dipole-dipole interaction in a degenerate quantum gas. *Phys. Rev. Lett.* **95**, 150406 (2005).
- Koch, T. *et al.* Stabilization of a purely dipolar quantum gas against collapse. *Nature Phys.* **4**, 218–222 (2008).
- Tang, Y., Sykes, A., Burdick, N. Q., Bohn, J. L. & Lev, B. L. s-wave scattering lengths of the strongly dipolar bosons <sup>162</sup>Dy and <sup>164</sup>Dy. *Phys. Rev. A* **92**, 022703 (2015).
- Maier, T. *et al.* Broad universal Feshbach resonances in the chaotic spectrum of dysprosium atoms. *Phys. Rev. A* **92**, 060702(R) (2015).
- Bradley, C. C., Sackett, C. A. & Hulet, R. G. Bose-Einstein condensation of lithium: observation of limited condensate number. *Phys. Rev. Lett.* **78**, 985–989 (1997).
- Lee, T. D., Huang, K. & Yang, C. N. Eigenvalues and eigenfunctions of a Bose system of hard spheres and its low-temperature properties. *Phys. Rev.* **106**, 1135–1145 (1957).
- Petrov, D. S. Quantum mechanical stabilization of a collapsing Bose-Bose mixture. *Phys. Rev. Lett.* **115**, 155302 (2015).
- Lima, A. R. P. & Pelster, A. Quantum fluctuations in dipolar Bose gases. *Phys. Rev. A* **84**, 041604 (2011).
- Saccani, S., Moroni, S. & Boninsegni, M. Excitation spectrum of a supersolid. *Phys. Rev. Lett.* **108**, 175301 (2012).

**Acknowledgements** We thank A. Griesmaier for support at the early stage of the experiment and D. Zajec, D. Peter, H. P. Büchler and L. Santos for discussions. This work is supported by the German Research Foundation (DFG) within SFB/TRR21. H.K. acknowledges support by the ‘Studienstiftung des deutschen Volkes’.

**Author Contributions** All authors discussed the results, made critical contributions to the work and contributed to the writing of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.K. ([h.kadau@physik.uni-stuttgart.de](mailto:h.kadau@physik.uni-stuttgart.de)) or T.P. ([t.pfau@physik.uni-stuttgart.de](mailto:t.pfau@physik.uni-stuttgart.de)).

## METHODS

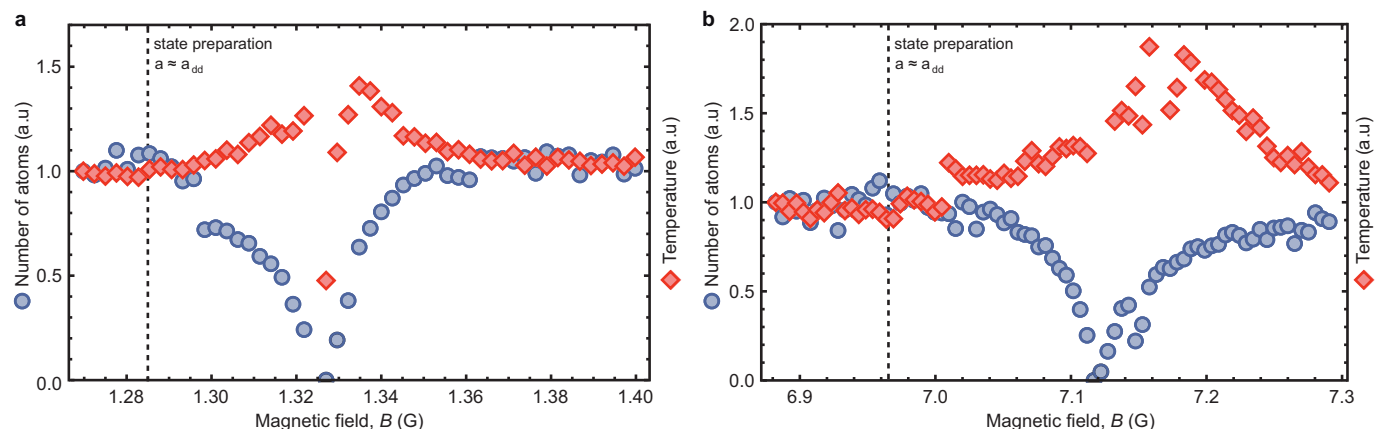
**Preparation of a Dy BEC.** Bosonic  $^{164}\text{Dy}$  atoms are first cooled in a narrow-line magneto-optical trap, operating on the 626-nm optical cycling transition, which polarizes the atoms to the lowest Zeeman state  $J=8$ ,  $m_J=-8$ , and subsequently directly loaded into a single-beam optical dipole trap (ODT)<sup>19</sup> created by a broadband fibre laser operating at 1,070 nm with a power of 72 W. By moving the focusing lens of the transport ODT over a range of 375 mm, we transport the atoms into a glass cell. After the transport, we load the atoms from the transport ODT into a crossed optical dipole trap created by a single-mode laser operating at 1,064 nm. Following this, the atoms are Doppler cooled with 626-nm light, and we utilize forced evaporative cooling by ramping down the power of both trapping laser beams at a magnetic field of  $B=1.012$  G far away from any resonance. When close to degeneracy, we tune the magnetic field close to a Feshbach resonance and achieve there a quasi-pure BEC with typically 15,000 atoms and a temperature of  $T=70$  nK. Before shaping the final trap with harmonic frequencies of  $(\nu_x, \nu_y, \nu_z)=(46, 44, 133)$  Hz, we apply an additional magnetic gradient of  $1.1\text{ G cm}^{-1}$  to partially compensate gravity.

**Feshbach resonance.** The background scattering length of  $^{164}\text{Dy}$  has been measured to be  $a_{\text{bg}}=92(8)a_0$ , in which  $a_0$  is the Bohr radius<sup>24,25</sup>. In the vicinity of a Feshbach resonance, the scattering length scales as  $a=a_{\text{bg}}(1-\Delta B/(B-B_0))$ , with  $B$  the applied magnetic field,  $B_0$  the centre and  $\Delta B$  the width of the Feshbach resonance. To tune the scattering length  $a$ , we used a narrow resonance at  $B_0=7.117(3)$  G with  $\Delta B=51(15)$  mG. We calibrated the magnetic field with a weak radiofrequency field driving transitions between magnetic sublevels. To measure the resonance position, we use atom trap-loss spectroscopy and find the maximal loss at the central position  $B_0$  of the Feshbach resonance. To estimate the width of the resonances we did a thermalization experiment at different magnetic

field positions and found a maximum temperature at the field position  $B_0 + \Delta B$  (Extended Data Fig. 1b). Close to the maximal temperature we observed another narrow resonance that influences our width measurement. Hence, we cannot state precise values for the scattering length. In addition, we did the same investigations for the structured states on a second, even narrower resonance located at  $B_0=1.326(3)$  G with a width of  $\Delta B=8(5)$  mG (Extended Data Fig. 1a). We observed the same qualitative results with this resonance, which indicates that self-organization is independent of a particular Feshbach resonance.

**In situ phase-contrast polarization imaging.** Phase-contrast polarization imaging was first introduced with lithium atoms<sup>26</sup> and relies on the dispersive phase shift instead of direct absorption giving rise to the optical density. We use an off-resonant beam ( $\Delta=-1.1\text{ GHz}=-35I$ ) near the strongest optical transition at 421 nm with a linewidth of  $\Gamma\approx 32$  MHz to suppress absorption. The imaging beam is linearly polarized and we apply a magnetic field of more than 1 G in the beam propagation direction so that the atoms see a mixture of left- and right-handed circularly polarized light. In our case, the Dy atoms are fully spin-polarized in the lowest-lying Zeeman state  $m_J=-8$ . Thus, the atoms couple mainly to the  $\sigma^-$  optical transition, and the  $\sigma^+$  transition is suppressed by a factor of about 150, owing to the difference in the Clebsch–Gordan coefficients. Hence, the atoms show a strong circular birefringence or optical rotation. If both coupled and uncoupled polarizations are combined on a linear polarizer, then the transferred intensity distribution depends on the dispersive shift of the atoms. The atomic plane is then imaged through a commercial diffraction-limited objective (numerical aperture of 0.32), corrected for the upper window of our glass cell, with a resolution of  $1\text{ }\mu\text{m}$ . With a second commercial objective, the image is magnified by a factor of 50 and guided through a polarizer to the camera.

**Sample size.** No statistical methods were used to predetermine sample size.



**Extended Data Figure 1 | Atom trap-loss spectroscopy for two Feshbach resonances.** Atom-loss spectroscopy mapping Feshbach resonances of  $^{164}\text{Dy}$ . The number of atoms (blue circles) and temperature (red diamonds) are normalized. **a**, The atom-number minimum shows the centre of the Feshbach resonance at  $B_0 = 1.326(3)$  G, while the temperature is maximized at  $B_0 + \Delta B$ , with  $\Delta B = 8(5)$  mG. We prepared stable BECs

at  $B = 1.285(3)$  G (dashed black line) before we induced an instability for lower magnetic-field values. **b**, For the data presented in the main text, we used a resonance at  $B_0 = 7.117(3)$  G (the atom-number minimum) with  $\Delta B = 51(15)$  mG. Stable BECs were created at  $B = 6.962(3)$  G (dashed black line).



# Observation of polar vortices in oxide superlattices

A. K. Yadav<sup>1,2\*</sup>, C. T. Nelson<sup>1,3,4\*</sup>, S. L. Hsu<sup>1,3,4</sup>, Z. Hong<sup>5</sup>, J. D. Clarkson<sup>1,3</sup>, C. M. Schlepütz<sup>6</sup>, A. R. Damodaran<sup>1</sup>, P. Shafer<sup>7</sup>, E. Arenholz<sup>7</sup>, L. R. Dedon<sup>1</sup>, D. Chen<sup>1,3</sup>, A. Vishwanath<sup>2,3</sup>, A. M. Minor<sup>1,2,4</sup>, L. Q. Chen<sup>5</sup>, J. F. Scott<sup>8</sup>, L. W. Martin<sup>1,2</sup> & R. Ramesh<sup>1,2,3</sup>

**The complex interplay of spin, charge, orbital and lattice degrees of freedom provides a plethora of exotic phases and physical phenomena<sup>1–5</sup>. In recent years, complex spin topologies have emerged as a consequence of the electronic band structure and the interplay between spin and spin–orbit coupling in materials<sup>6,7</sup>. Here we produce complex topologies of electrical polarization—namely, nanometre-scale vortex–antivortex (that is, clockwise–anticlockwise) arrays that are reminiscent of rotational spin topologies<sup>6</sup>—by making use of the competition between charge, orbital and lattice degrees of freedom in superlattices of alternating lead titanate and strontium titanate layers. Atomic-scale mapping of the polar atomic displacements by scanning transmission electron microscopy reveals the presence of long-range ordered vortex–antivortex arrays that exhibit nearly continuous polarization rotation. Phase-field modelling confirms that the vortex array is the low-energy state for a range of superlattice periods. Within this range, the large gradient energy from the vortex structure is counterbalanced by the corresponding large reduction in overall electrostatic energy (which would otherwise arise from polar discontinuities at the lead titanate/strontium titanate interfaces) and the elastic energy associated with epitaxial constraints and domain formation. These observations have implications for the creation of new states of matter (such as dipolar skyrmions, hedgehog states) and associated phenomena in ferroic materials, such as electrically controllable chirality.**

The ability to synthesize heteroepitaxial complex oxides has enabled unprecedented access to high-quality, single-crystalline materials and routes to manipulate order parameters<sup>8,9</sup>. In ferroelectric materials (that is, those possessing a spontaneous, switchable electrical polarization), epitaxial strain and advances in layer-by-layer growth techniques have enabled the study of the fundamental limits of ferroelectricity<sup>10,11</sup> and the discovery of novel interfacial phenomena<sup>12,13</sup>. In short-period (of only a few unit cells) superlattices of PbTiO<sub>3</sub>/SrTiO<sub>3</sub>, researchers observed the emergence of ‘improper’ ferroelectricity arising from octahedral tilts in the SrTiO<sub>3</sub> layer<sup>14–16</sup> and this has motivated a number of additional studies of such effects<sup>17</sup>. At relatively large length scales (greater than 5 nm), interfaces still play a formative role, driving both bulk<sup>18</sup> and thin-film<sup>19–21</sup> ferroelectrics to form patterns of flux-closure polar domains, whose microstructures have been the topic of extensive research including atomic-scale polarization studies<sup>19,22,23</sup>. At intermediate length scales (tens of unit cells), theoretical studies have suggested that there is potential for topological structures such as vortices, waves and skyrmions<sup>24–30</sup>, depending on the interplay between strain, depolarization and gradient energies. In this work, we demonstrate ordered arrays of polar vortices, made up of vortex–antivortex pairs, in ferroelectric/paraelectric superlattices.

Symmetric (SrTiO<sub>3</sub>)<sub>n</sub>/(PbTiO<sub>3</sub>)<sub>n</sub> superlattices with  $n = 2–27$  were synthesized on DyScO<sub>3</sub> (001)<sub>pc</sub> (where pc refers to the pseudocubic

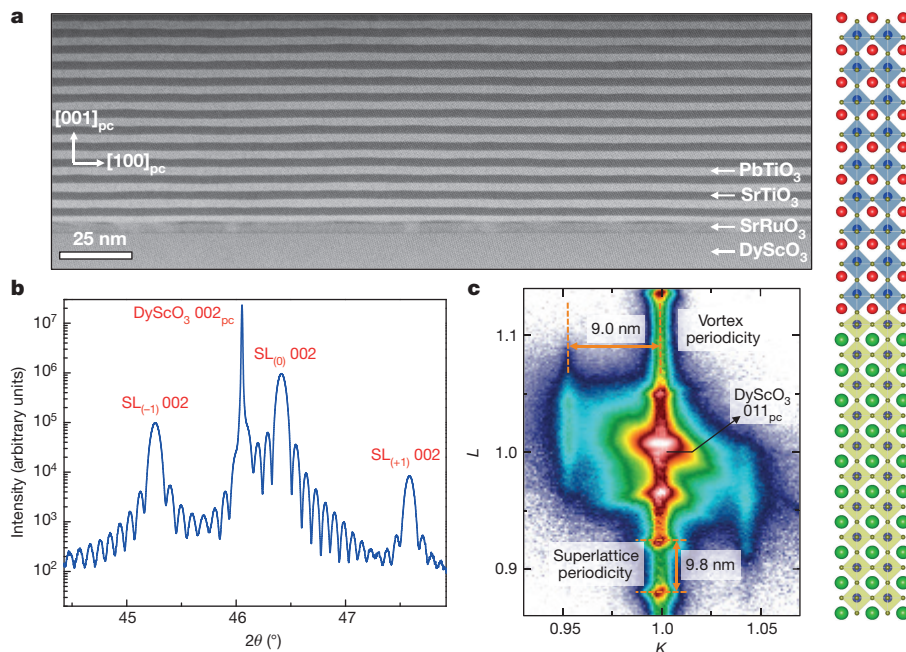
notation) substrates via reflection high-energy electron diffraction (RHEED)-assisted pulsed-laser deposition (details of the growth are provided in Methods and Extended Data Fig. 1). Superlattices are henceforth referred to using the ‘ $n \times n$ ’ shorthand wherein  $n$  corresponds to the thickness of the SrTiO<sub>3</sub> and PbTiO<sub>3</sub> layers in unit cells (structures depicted schematically on the right of Fig. 1a). Chemical analysis via Rutherford backscattering spectrometry (RBS) confirms stoichiometric PbTiO<sub>3</sub> within the detectability limits of the technique ( $\pm 1\%$ , Extended Data Fig. 2). A typical low-magnification, scanning transmission electron microscopy (STEM, details in Methods) image of the cross-section of a  $10 \times 10$  superlattice taken along the [010]<sub>pc</sub> zone axis reveals the layer uniformity (Fig. 1a), and atomic-scale high-resolution STEM (HR-STEM) confirms sharp and coherent interfaces (Extended Data Fig. 3a). X-ray diffraction studies including symmetric Bragg scans (Fig. 1b) and reciprocal space maps (RSMs, Fig. 1c) reveal superlattice reflections in the out-of-plane direction ( $L$ , [001]<sub>pc</sub>) corresponding to the superlattice period of 9–10 nm. Side-lobe diffraction peaks are also observed along the in-plane direction ( $K$ , [010]<sub>pc</sub>) corresponding to a 9–10 nm periodicity (arising from a polar ordering which is discussed in the following section).

Mapping of the atomic polar displacement ( $P_{PD}$ ) was performed to determine the polarization distribution within the superlattices. A displacement vector-mapping algorithm<sup>23</sup> (details in Methods and Extended Data Figs 3 and 4) was implemented on the cross-sectional HR-STEM images to measure local non-centrosymmetry of the lattice. A vector map of these polar displacements within a  $10 \times 10$  superlattice (Fig. 2a) shows the formation of long-range, ordered arrays of vortex structures with alternating rotation directions. The lateral periodicity is approximately the same as the superlattice period ( $\sim 10$  nm), closely matching the in-plane ( $K$ , [010]<sub>pc</sub>) periodicity observed in the RSMs. These vortices exhibit a continuous rotation of the local polarization vector, in contrast to segregation into uniform domains and domain walls typical of ferroelectric domain and flux-closure structures<sup>19–21</sup>. As a consequence, a dominant fraction of the PbTiO<sub>3</sub> layers exhibit a continual rotation of  $P_{PD}$ , as illustrated in maps of the curl (that is, vorticity) of the displacement vector field ( $\nabla \times P_{PD}$ )<sub>[010]</sub> (Fig. 2c; additional details are provided in Methods and the full curl map corresponding to Fig. 2a is found in Extended Data Fig. 5a). To better understand the energetics of the vortex–antivortex states, we carried out phase-field modelling of the polarization structure as a function of superlattice periodicity (details in Methods). The phase-field-calculated polarization maps for the same  $10 \times 10$  superlattice indicate the formation of a vortex–antivortex ground state bearing close resemblance to the experimental observations (Fig. 2d).

Both in- and out-of-plane long-range vortex ordering is observed by electron diffraction and dark-field transmission electron microscopy (DF-TEM) along cross-section and planar views (Fig. 3a and b, respectively). The cross-sectional DF-TEM images show a long-range

<sup>1</sup>Department of Materials Science and Engineering, University of California, Berkeley, California 94720, USA. <sup>2</sup>Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. <sup>3</sup>Department of Physics, University of California, Berkeley, California 94720, USA. <sup>4</sup>National Center for Electron Microscopy, Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. <sup>5</sup>Department of Materials Science and Engineering, Pennsylvania State University, State College, Pennsylvania 16802, USA. <sup>6</sup>Advanced Photon Source, Argonne National Laboratory, Argonne, Illinois 60439, USA. <sup>7</sup>Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. <sup>8</sup>Schools of Chemistry and Physics, University of St Andrews, St Andrews KY16 9ST, UK.

\*These authors contributed equally to this work.



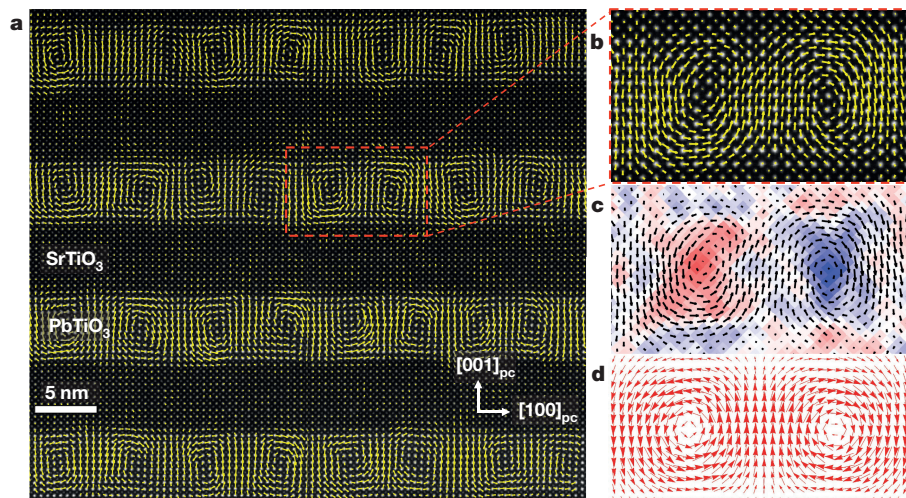
**Figure 1 | Structural characterization of  $(\text{SrTiO}_3)_n/(\text{PbTiO}_3)_n$  superlattices.** **a**, Low-magnification, STEM image of the cross-section of a  $(\text{SrTiO}_3)_{10}/(\text{PbTiO}_3)_{10}$  superlattice, together with a schematic illustration of the superlattice structure (right). The  $\text{PbTiO}_3$  (top) is illustrated for the case of a  $[001]_{\text{pc}}$  (upward) polarization which results from the downward displacement of the Ti-'centred' oxygen octahedra relative to Pb (red, blue, green and gold spheres indicate Pb, Ti, Sr and O atoms, respectively; blue- and green-shaded regions surrounding Ti atoms show oxygen octahedra). **b**, High-resolution  $\theta$ - $2\theta$  symmetric scan of a 100-nm  $(\text{SrTiO}_3)_{10}/(\text{PbTiO}_3)_{10}$  superlattice showing the presence of Pendellösung fringes, which attest to the smoothness of the surface and the high quality of the 100-nm-thick superlattices (see Methods for details and nomenclature used for peak labels). **c**, Synchrotron-based, in-plane X-ray RSM study of a  $(\text{SrTiO}_3)_{12}/(\text{PbTiO}_3)_{12}$  superlattice about the  $\text{DyScO}_3$   $011_{\text{pc}}$  diffraction condition. The asymmetric  $L$ - $K$  scan reveals satellite peaks corresponding to vortex-antivortex pairs with  $\sim 10$  nm periodicity.

periodic array of intensity modulation which corresponds to the vortex-antivortex pairs with a spacing of 9–10 nm both in and out of the plane of the superlattice. Like the X-ray RSM studies, the long-range order also manifests as additional superlattice reflections in the selected-area electron diffraction (SAED) patterns in both cross-section and planar views (see the magnified  $001_{\text{pc}}$  and  $050_{\text{pc}}$  reflections from the SAED patterns in Fig. 3a inset and Fig. 3b inset, respectively). The uniform array of stripes in the planar-view DF-TEM (Fig. 3b), again with a 9–10 nm periodicity, suggests that the vortices extend over distances of at least  $1\ \mu\text{m}$ . We also note that vortices are observed to exist in both in-plane directions (Extended Data Fig. 6).

Phase-field modelling of the  $\text{PbTiO}_3/\text{SrTiO}_3$  superlattices was performed to study the energetics and to ascertain the 3D structure of the vortex-antivortex pairs. A range of symmetric superlattice periodicities were simulated, allowing each to evolve from a random initial state according to the time-dependent Ginzburg–Landau equation (details in Methods). The 3D structure of the  $10 \times 10$  superlattice, corresponding to our experimental HR-STEM studies, is shown in Fig. 4. With no a priori assumption of the polarization ( $\mathbf{P}$ ) distribution, the model converges to a vortex-antivortex array structure. For

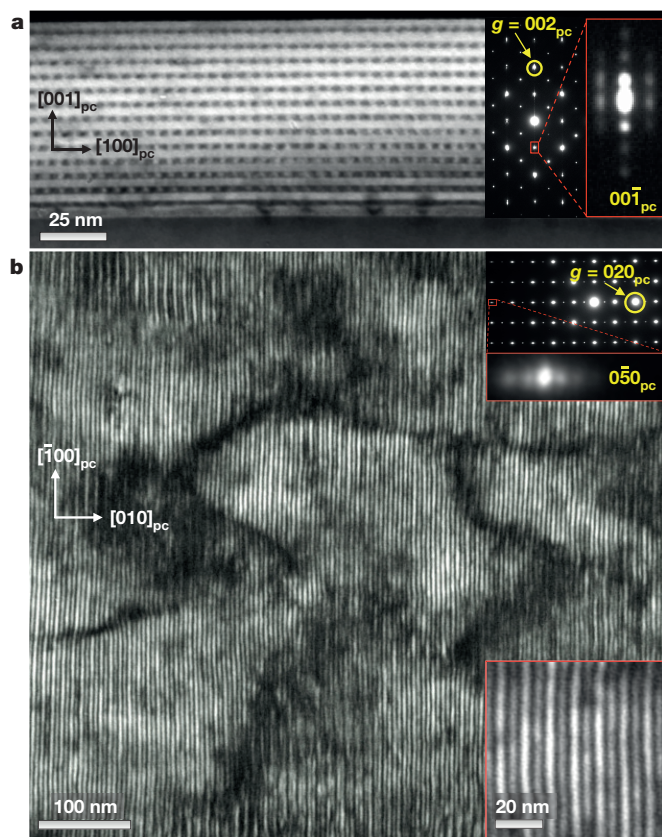
comparison, sub-regions of the HR-STEM and planar-view DF-TEM images are projected on the same 3D axes as the model (Fig. 4). The tubular, ordered vortices match the experimentally observed structures closely. The polarization rotation extracted from the model is mapped onto the front plane and in 3D isosurfaces (Fig. 4) using a blue/red colour scale corresponding to alternating clockwise (vortices) and anticlockwise (antivortices) according to the  $(\nabla \times \mathbf{P})_{[010]}$  curl of the polarization. Two additional slices through the 3D structure show the component of polarization along  $[100]_{\text{pc}}$  ( $\mathbf{P}_{[100]}$ ) and  $[001]_{\text{pc}}$  ( $\mathbf{P}_{[001]}$ ). Like the polar displacements from the HR-STEM, the model lacks discrete domains and, instead, exhibits a smooth rotation of the polarization. The long-range order is also in good agreement, preferring interlayer alignment of the vortices, particularly with matching rotation directions (Extended Data Fig. 5c).

On the basis of both the experimental results and the phase-field simulations, the observed vortex structure results (primarily) from competition between three energies. The first is the elastic energy, which arises from the fact that the  $\text{PbTiO}_3$  is epitaxially constrained (in a tensile fashion) by the  $\text{DyScO}_3$  substrate which traditionally drives the  $\text{PbTiO}_3$



**Figure 2 | Observation of vortex-antivortex structures.** **a**, Cross-sectional HR-STEM image with an overlay of the polar displacement vectors ( $\mathbf{P}_{\text{PD}}$ , indicated by yellow arrows) for a  $(\text{SrTiO}_3)_{10}/(\text{PbTiO}_3)_{10}$  superlattice, showing that an array of vortex-antivortex pairs is present in each  $\text{PbTiO}_3$  layer. **b**, A magnified image of a single vortex-antivortex pair, showing the full density of data points (one for each atom) and the continuous rotation of the polarization state within such vortex-antivortex pairs. **c**, The curl of the polar displacement  $(\nabla \times \mathbf{P}_{\text{PD}})_{[010]}$  for the same vortex-antivortex pair reveals the alternating rotation directions of the structures (see details in Methods and Extended Data Fig. 5a and b). The  $(\nabla \times \mathbf{P}_{\text{PD}})_{[010]}$  curl value is plotted with a red/blue colour scale where no-vorticity (curl = 0) is white, clockwise (negative) is blue and anticlockwise (positive) is red. **d**, Polarization vectors from a phase-field simulation of the same  $(\text{SrTiO}_3)_{10}/(\text{PbTiO}_3)_{10}$  superlattice, which predicts vortex-antivortex pairs that closely match the experimental observations.



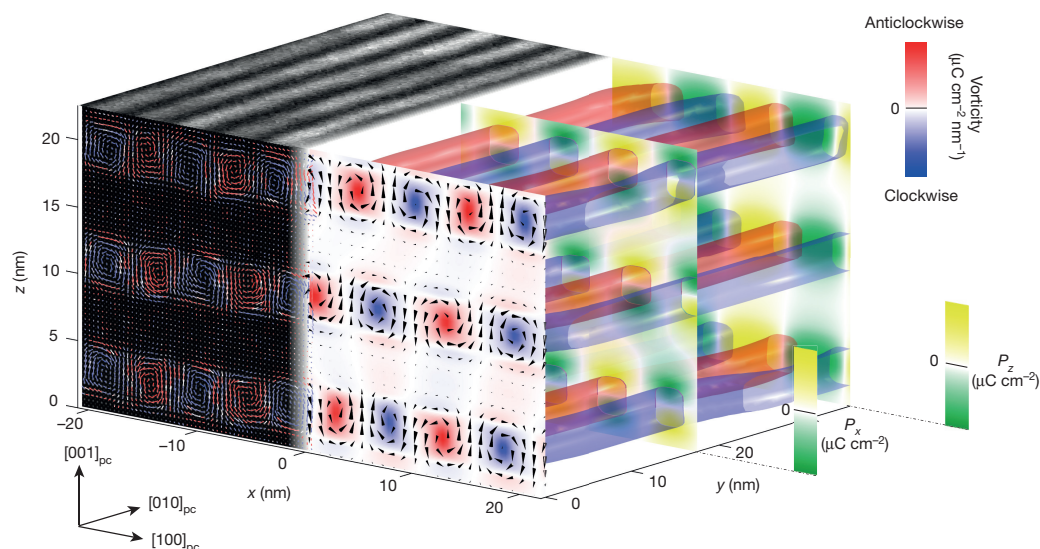


**Figure 3 | Long-range ordering of vortices.** **a**, Main panel, cross-sectional DF-TEM images of a  $(\text{SrTiO}_3)_{10}/(\text{PbTiO}_3)_{10}$  superlattice revealing long-range ordering of vortices (each bright/dark modulation corresponds to a period of the vortex–antivortex structure). Inset, SAED pattern for the sample with the  $g$ -vector of the two-beam imaging condition indicated by the yellow arrow. Close inspection of the SAED pattern reveals superlattice reflections both in and out of the plane, suggesting long-range ordering of the vortex structures (boxed area of inset with  $00\bar{1}_{\text{pc}}$  reflection is shown magnified at right). **b**, Planar-view DF-TEM imaging of a  $(\text{SrTiO}_3)_{16}/(\text{PbTiO}_3)_{16}$  superlattice which likewise exhibits long-range in-plane ordering associated with the vortex–antivortex pairs along  $[\bar{1}00]_{\text{pc}}$  and confirms the widespread occurrence of these vortex pairs in the superlattice films. The upper inset shows the SAED pattern for the sample with the  $g$ -vector of the two-beam imaging condition indicated by the yellow arrow. Again, close inspection of the diffraction pattern (boxed area of inset shown magnified) reveals in-plane superlattice reflections indicating long-range ordering of the vortex structures. The lower inset shows a magnified view of the vortex structures, again revealing  $\sim 10$  nm periodicities.

to adopt a mixture of in-plane  $P_{[100]/[010]}$  and out-of-plane  $P_{[001]}$  polarizations (that is, a  $c/a$  domain structure). The second major contribution is electrostatic energy from built-in electric fields. As one transitions from the  $\text{PbTiO}_3$  to the  $\text{SrTiO}_3$  layer there is likely to be a large polar discontinuity normal to the interface and thus there is a strong driving force for bound charges since  $\nabla \cdot \mathbf{P} \neq 0$ . Finally, the third major energy in competition is the gradient energy. This is the energy required to rotate or change the direction or magnitude of the polarization. All told,

in this system, these three energies combine to create a new topology of polarization—the system localizes the in-plane  $P_{[100]/[010]}$  components at the  $\text{PbTiO}_3/\text{SrTiO}_3$  interfaces and forms alternating out-of-plane  $P_{[001]}$  components mid-layer—thereby creating a rotating polarization.

It should also be noted that the balance of these terms, and thus the ground-state structure, is strongly dependent on the superlattice period. In short-period superlattices ( $n < 10$ ), the vortex becomes unstable and, instead, the  $\text{PbTiO}_3$  minimizes the electrostatic and gradient energy at the expense of elastic energy by adopting only in-plane  $P_{[100]/[010]}$  components (that is,  $a_1/a_2$  domain structures). As the superlattice period increases to intermediate length scales, a discontinuity in the total energy appears in the phase diagram (derived from the phase-field simulations and total energy calculations) and thus a co-existence of the  $a_1/a_2$  and vortex–antivortex structures is observed. In long-period superlattices ( $n \gtrsim 18$ ), the system begins to evolve into classical flux-closure domain structures characterized by the formation of discrete domains surrounding localized polarization gradients at domain walls. A side-by-side comparison of HR-STEM polar displacement maps and phase-field



**Figure 4 | Three-dimensional phase-field simulation for a  $(\text{SrTiO}_3)_{10}/(\text{PbTiO}_3)_{10}$  superlattice.** The 3D geometry of the vortex–antivortex array from phase-field simulation of a  $(\text{SrTiO}_3)_{10}/(\text{PbTiO}_3)_{10}$  superlattice is shown on the right. The front cross-section of the model shows the polarization vector  $\mathbf{P}$  ordered into clockwise (blue) and anticlockwise (red) vortex–antivortex states which extend along the  $[010]_{\text{pc}}$ . For comparison, on the left is shown a cross-sectional HR-STEM image overlaid with

a polar displacement vector map and a planar-view DF-TEM image projected onto the front and top planes of the axes, respectively. Red/blue colour scales correspond to the curl of the polarization extracted from the phase-field model and the HR-STEM polar displacement map ( $P_{\text{PD}}$ ). The scale bars on  $P_x$  and  $P_z$  are from  $-44.81$  to  $44.81$  ( $\mu\text{C cm}^{-2}$ ) and from  $-54.24$  to  $54.24$  ( $\mu\text{C cm}^{-2}$ ), respectively, whereas the vorticity is from  $-129.36$  to  $129.36$  ( $\mu\text{C cm}^{-2} \text{ nm}^{-1}$ ).



simulations for vortex ( $10 \times 10$ ) and flux-closure ( $50 \times 50$ ) structures is provided (Extended Data Fig. 7); we also show the superlattice-period-dependent evolution of structures from phase-field models (Extended Data Fig. 8). In summary, the continuous curling of polarization allows the simultaneous achievement of electrostatically favoured, non-diverging polarization and a strain-favoured mixture of in-plane and out-of-plane components of polarization, stabilizing novel vortex states.

These observations beget several important questions and avenues for future exploration. First, the vortices are observed to be slightly asymmetric (the cores not being perfectly centred and equidistant, Fig. 2a), which could be a source of chirality and potentially used to produce tunable optical behaviour. Second, although the long-range, in-plane order of the vortex structures is quite robust, the out-of-plane order is somewhat less so. This out-of-plane ordering is likely to be a low-energy excitation (akin to ferroelectric liquid crystals and Abrikosov lattices in superconductors) and consequently to exhibit a certain degree of disorder and the potential for multiple ground states, which may be a rich area for exploration of the macro-scale ordering by tuning the interlayer stacking. Last, the generation of vortex topologies could be extended to other material systems and other symmetries—for example, a multiferroic BiFeO<sub>3</sub> system could also provide the possibility of enhancing the canted moment arising from Dzyaloshinski–Moriya interactions.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 17 June; accepted 16 November 2015.**

**Published online 27 January 2016.**

- Tokura, Y. & Nagaosa, N. Orbital physics in transition-metal oxides. *Science* **288**, 462–468 (2000).
- Imada, M., Fujimori, A. & Tokura, Y. Metal-insulator transitions. *Rev. Mod. Phys.* **70**, 1039–1263 (1998).
- Zubko, P. et al. Interface physics in complex oxide heterostructures. *Annu. Rev. Condens. Matter Phys.* **2**, 141–165 (2011).
- Millis, A. J. Lattice effects in magnetoresistive manganese perovskites. *Nature* **392**, 147–150 (1998).
- Kivelson, S. A., Fradkin, E. & Emery, V. J. Electronic liquid-crystal phases of a doped Mott insulator. *Nature* **393**, 550–553 (1998).
- Nagaosa, N. & Tokura, Y. Topological properties and dynamics of magnetic skyrmions. *Nature Nanotechnol.* **8**, 899–911 (2013).
- Schulz, T. et al. Emergent electrodynamics of skyrmions in a chiral magnet. *Nature Phys.* **8**, 301–304 (2012).
- Schlom, D. G. et al. Strain tuning of ferroelectric thin films. *Annu. Rev. Mater. Res.* **37**, 589–626 (2007).
- Mannhart, J. & Schlom, D. G. Oxide interfaces — an opportunity for electronics. *Science* **327**, 1607–1611 (2010).
- Dawber, M., Rabe, K. M. & Scott, J. F. Physics of thin-film ferroelectric oxides. *Rev. Mod. Phys.* **77**, 1083–1130 (2005).
- Fong, D. D. et al. Ferroelectricity in ultrathin perovskite films. *Science* **304**, 1650–1653 (2004).
- Chakhalian, J., Millis, A. J. & Rondinelli, J. Whither the oxide interface. *Nature Mater.* **11**, 92–94 (2012).
- Yu, P. Interface control of bulk ferroelectric polarization. *Proc. Natl Acad. Sci. USA* **109**, 9710–9715 (2012).
- Bousquet, E. et al. Improper ferroelectricity in perovskite oxide artificial superlattices. *Nature* **452**, 732–736 (2008).
- Sichuga, D. et al. Chiral patterns of tilting of oxygen octahedra in zero-dimensional ferroelectrics and multiferroics: a first principle-based study. *Phys. Rev. Lett.* **104**, 207603 (2010).
- Zhao, H. J. et al. Atomistic theory of hybrid improper ferroelectricity in perovskites. *Phys. Rev. B* **89**, 174101(R) (2014).
- Pitcher, M. J. et al. Tilt engineering of spontaneous polarization and magnetization above 300K in a bulk layered perovskite. *Science* **347**, 420–424 (2015).
- McQuaid, R. G. P., McGilly, L. J., Sharma, P., Gruverman, A. & Gregg, J. M. Mesoscale flux-closure domain formation in single-crystal BaTiO<sub>3</sub>. *Nature Commun.* **2**, 404 (2011).
- Tang, Y. L. et al. Observation of periodic array of flux-closure quadrants in strained ferroelectric PbTiO<sub>3</sub> films. *Science* **348**, 547–551 (2015).
- Gruverman, A. et al. Vortex ferroelectric domains. *J. Phys. Condens. Matter* **20**, 342201 (2008).
- Balke, N. et al. Enhanced electric conductivity at ferroelectric vortex cores in BiFeO<sub>3</sub>. *Nature Phys.* **8**, 81–88 (2012).
- Jia, C.-L., Urban, K. W., Alexe, M., Hesse, D. & Vrejoiu, I. Direct observation of continuous electric dipole rotation in flux-closure domains in ferroelectric Pb(Zr, Ti)O<sub>3</sub>. *Science* **331**, 1420–1423 (2011).
- Nelson, C. T. et al. Spontaneous vortex nanodomain arrays at ferroelectric heterointerfaces. *Nano Lett.* **11**, 828–834 (2011).
- Naumov, I. I., Bellaiche, L. & Fu, H. Unusual phase transitions in ferroelectric nanodisks and nanorods. *Nature* **432**, 737–740 (2004).
- Choudhury, N., Walizer, L., Lisenkov, S. & Bellaiche, L. Geometric frustration in compositionally modulated ferroelectrics. *Nature* **470**, 513–517 (2011).
- Ponomareva, I., Naumov, I. & Bellaiche, L. Low-dimensional ferroelectrics under different electrical and mechanical boundary conditions: atomistic simulations. *Phys. Rev. B* **72**, 214118 (2005).
- Prosandeev, S. & Bellaiche, L. Characteristics and signatures of dipole vortices in ferroelectric nanodots: first-principles-based simulations and analytical expressions. *Phys. Rev. B* **75**, 094102 (2007).
- Prosandeev, S., Ponomareva, I., Naumov, I., Kornev, I. & Bellaiche, L. Original properties of dipole vortices in zero-dimensional ferroelectrics. *J. Phys. Condens. Matter* **20**, 193201 (2008).
- Sichuga, D. & Bellaiche, L. Epitaxial Pb(ZrTi)O<sub>3</sub> ultrathin films under open circuit electrical boundary conditions. *Phys. Rev. Lett.* **106**, 196102 (2011).
- Levanyuk, A. P. & Blinc, R. Ferroelectric phase transitions in small particles and local regions. *Phys. Rev. Lett.* **111**, 097601 (2013).

**Acknowledgements** A.K.Y. and R.R. acknowledge support from the Office of Basic Energy Sciences, US Department of Energy (DE-AC02-05CH11231), for the synthesis and characterization of superlattices. R.R. and C.T.N. acknowledge support from the Office of Basic Energy Sciences, US Department of Energy (DE-AC02-05CH11231), for TEM characterization of superlattices. S.L.H. acknowledges support from the National Science Foundation under the MRSEC program (DMR-1420620). Z.H. acknowledges support from the National Science Foundation (DMR-1210588). J.D.C. acknowledges support from the Office of Basic Energy Sciences, US Department of Energy (DE-AC02-05CH11231). C.M.S. acknowledges use of the Advanced Photon Source, which was supported by the US Department of Energy, Office of Science, Office of Basic Energy Science (DE-AC02-06CH11357), for the synchrotron-based reciprocal space map studies of superlattice structures at Sector 33-BM-C beamline. A.R.D. acknowledges support from the Army Research Office (W911NF-14-1-0104). P.S. and E.A. acknowledge support from the Director, Office of Science, Office of Basic Energy Sciences, US Department of Energy (DE-AC02-05CH11231), for the synchrotron-based studies of the superlattice structures. L.R.D. acknowledges support from the US Department of Energy, Office of Basic Energy Sciences (DE-SC0012375) for the chemical analysis of the superlattice structures. D.C. acknowledges support from the National Science Foundation under the MRSEC program (DMR-1420620). A.V. and A.M.M. acknowledge support from Office of Basic Energy Sciences, US Department of Energy (DE-AC02-05CH11231). L.Q.C. acknowledges support from the National Science Foundation (DMR-1210588). L.W.M. acknowledges support from the National Science Foundation (DMR-1451219). Electron microscopy of superlattice structures was performed at the Molecular Foundry, LBNL, supported by the Office of Science, Office of Basic Energy Sciences, US Department of Energy (DE-AC02-05CH11231). Z.H. thanks J. Britson, F. Xue and J. Wang for help and discussions.

**Author Contributions** A.K.Y. and R.R. designed the experiments. A.K.Y. carried out the synthesis and characterization of superlattice samples. C.T.N. and S.L.H. performed TEM characterization of superlattice samples, along with the detailed polarization vector analysis. Z.H. performed phase-field calculations for these superlattices. C.M.S. did the reciprocal space map studies of superlattice structures using synchrotron X-ray diffraction. L.R.D. did the chemical analysis of the superlattice films using Rutherford backscattering. D.C., J.D.C. and A.R.D. did the piezoforce microscopy studies of superlattice films. P.S. and E.A. did synchrotron X-ray circular dichroism measurements on superlattice films. R.R., L.W.M., A.K.Y., C.T.N., A.R.D., J.D.C., Z.H., L.Q.C., A.V. and J.F.S. analysed the data and co-wrote the manuscript. R.R., L.W.M., A.M.M. and L.Q.C. supervised the research. All authors contributed to the discussions and manuscript preparation.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.R. (rramesh@berkeley.edu).

## METHODS

**Material system.** The materials used in this study are lead titanate (PbTiO<sub>3</sub>), strontium titanate (SrTiO<sub>3</sub>), strontium ruthenate (SrRuO<sub>3</sub>) and dysprosium scandate (DyScO<sub>3</sub>). PbTiO<sub>3</sub> has a tetragonal structure with space group *P4mm* and bulk lattice parameters<sup>31</sup>  $a = 3.904$  Å and  $c = 4.178$  Å. PbTiO<sub>3</sub> is a ferroelectric at room temperature with an axis of spontaneous polarization along [001] and undergoes a phase transition<sup>31</sup> from a tetragonal (ferroelectric) to a cubic (paraelectric) structure at 490 °C in the bulk. SrTiO<sub>3</sub>, on the other hand, is a band insulator at room temperature with a cubic structure having a bulk lattice parameter of  $a = 3.905$  Å and space group *Pm3m*. SrRuO<sub>3</sub> is a ferromagnetic metal at room temperature with an orthorhombic structure having space group *Pbnm* and bulk lattice parameters<sup>32</sup> of  $a = 5.567$  Å,  $b = 5.530$  Å and  $c = 7.845$  Å. DyScO<sub>3</sub> is a band insulator at room temperature and has an orthorhombic structure with space group *Pbnm* and bulk lattice parameters<sup>33</sup> of  $a = 5.440$  Å,  $b = 5.717$  Å and  $c = 7.903$  Å. The orientation between the pseudocubic indices used in the text and the orthorhombic indices of the DyScO<sub>3</sub> substrate are (001)<sub>pc</sub> || (110)<sub>o</sub>, (100)<sub>pc</sub> || (110)<sub>o</sub>, (010)<sub>pc</sub> || (001)<sub>o</sub>. **Synthesis of the superlattices using RHEED-assisted pulsed-laser deposition.** Superlattices of (SrTiO<sub>3</sub>)<sub>*n*</sub>/(PbTiO<sub>3</sub>)<sub>*n*</sub> were synthesized on 5 nm SrRuO<sub>3</sub>-buffered, single crystalline DyScO<sub>3</sub> (001)<sub>pc</sub> substrates via pulsed-laser deposition. The growth temperature and oxygen pressure for the growth of the superlattices was 630 °C and 100 mtorr, respectively, to ensure stoichiometric transfer of both PbTiO<sub>3</sub> and SrTiO<sub>3</sub>. Growth of the SrRuO<sub>3</sub> was accomplished at a growth temperature of 700 °C and an oxygen pressure of 50 mtorr. All depositions were performed at a laser fluence of  $1.5 \text{ J cm}^{-2}$ . The growth was monitored using reflection high-energy electron diffraction (RHEED) to ensure maintenance of a layer-by-layer growth mode throughout the growth process (Extended Data Fig. 1) which, in turn, enables controlled growth of the superlattices with different period-thicknesses while achieving the same total thickness. The RHEED oscillations were monitored for specular spots along [100]<sub>pc</sub> of the DyScO<sub>3</sub> (001)<sub>pc</sub> substrate. After the growth, the samples were cooled to room temperature in 50 torr of oxygen to promote full oxidation. The use of a Pb<sub>1.2</sub>TiO<sub>3</sub> target was found to be critical in enabling persistent layer-by-layer growth (evidenced by RHEED oscillations obtained throughout the entire growth process, Extended Data Fig. 1), whereas the use of a stoichiometric PbTiO<sub>3</sub> target resulted in rapid transition to 3D (island) growth.

**Structural characterization. Laboratory-based X-ray diffraction.** Superlattices were characterized with a Panalytical X'Pert Pro X-ray Diffraction (XRD) machine with a Cu source. Symmetric XRD scans (Fig. 1b) were extensively performed to fully characterize the structural order in the out-of-plane direction, [001]<sub>pc</sub>, in these superlattice structures. The superlattice period achieved via RHEED-controlled growth was later confirmed using high-resolution  $\theta$ -2 $\theta$  symmetric XRD scans (see, for example, Fig. 1b). The presence of thickness fringes (*Pendellösung* fringes) around both the central Bragg peak (SL<sub>(0)</sub> 002) and the superlattice peaks (SL<sub>(-1)</sub> 002 and SL<sub>(+1)</sub> 002) are clearly visible in  $10 \times 10$  superlattices (Fig. 1b) and attest to the high crystallinity of the films, the precise nature of the interfaces, and the overall smoothness of these structures.

**Synchrotron X-ray diffraction.** This was employed to study the structure (both in and out of the plane) of the superlattices. The diffraction studies were performed at the Sector 33-BM-C beamline of the Advanced Photon Source, Argonne National Laboratory. The high flux from a synchrotron source delivered at this beamline was optimal for extracting lattice modulations associated with these vortex-antivortex pairs. In order to obtain a highly monochromatic beam with negligible higher-order harmonics, a double crystal monochromator in conjunction with two mirrors was used. Moreover, excellent accuracy of a Huber 4-circle diffractometer in combination with a PILATUS 100K pixel detector allowed us to determine the orientation of the crystal reliably and obtain 3D RSMs with high accuracy. The X-ray wavelength used for obtaining the RSMs was 0.8 Å ( $E = 15.5$  keV).

**Chemical analysis by RBS.** This was employed to confirm the stoichiometry of the PbTiO<sub>3</sub> at the optimized growth conditions and to eliminate any spurious effects of non-stoichiometry on the observation of these vortex structures. The results for a 200 nm PbTiO<sub>3</sub> film are provided (Extended Data Fig. 2). These results and the corresponding fit to the data confirm the chemical composition and homogeneity of the films within the limits of experimental errors. The spectra were taken with a 3.04 MeV He ion beam incident at an angle of 22.5° relative to the sample normal. The backscattered He ions were collected by a silicon surface-barrier detector positioned at 168° with respect to the incident beam. The SIMNRA software package was used to simulate the RBS spectra and obtain the composition of the film.  $R^2$  analysis was performed about the Pb and Ti peaks to avoid artificially increasing the value of  $R^2$  with the inclusion of substrate peaks.  $R^2$  was calculated using  $R^2 = 1 - \frac{\sum (e_i - f_i)^2}{\sum (e_i - \bar{e})^2}$ , where  $e_i$  and  $f_i$  values correspond to experimental and simulated data points, respectively, for different channel energies (indicated by variable  $i$ ), and  $\bar{e}$  indicates the mean of experimental data points.

**Conventional and scanning transmission electron microscopy.** TEM samples were prepared by mechanical polishing on an Allied High Tech Multiprep. Cross-sectional samples were polished at a 0.5° wedge, and planar samples with a flat polish. Samples were subsequently Ar ion milled using a Gatan Precision Ion Milling System with starting energies of 4 keV stepped down to a final cleaning energy of 200 eV.

Diffraction contrast TEM (Fig. 3) was performed on a JEOL 3010 at 300 kV. Dark field imaging was performed in the two-beam condition, the diffracted beam travelling along the optical axis of the TEM. The samples were tilted off zone-axis to isolate selected crystallographic planes used for imaging:  $g = 002_{pc}$  for Fig. 3a and Extended Data Fig. 6a;  $g = 020_{pc}$  for Figs 3b and 4; and  $g = 00\bar{2}_{pc}$  for Extended Data Fig. 6b. The tilt condition used for Fig. 3a and Extended Data Fig. 6b is shown by the SAED pattern (inset) in the latter. SAED patterns inset in Fig. 3a and b were collected along the zone axis.

Scanning TEM was performed on a Cs-corrected TEAM0.5 FEI Titan microscope at 300 kV using a high-angle annular detector resulting in 'Z-contrast' images. Raster distortion was minimized using a pair of HR-STEM images of the same region taken with orthogonal scan orientations. These were used to calculate a corrected 'pre-distortion' image to minimize the slow scan axis errors<sup>34</sup>.

Determination and mapping of the polar atomic displacements ( $P_{PD}$ ) was performed on the corrected HR-STEM HAADF images (Extended Data Fig. 3a, b) using local A- and B-site sublattice offset measurements<sup>23</sup>. The rationale is that for many displacive perovskite ferroelectrics the electric dipole moment is manifested by relative shifts of the cations, as is the case for PbTiO<sub>3</sub>, and this offset can be used to infer the full polarization ( $P$ ). This is very helpful since the anions (oxygen) needed to fully characterize an electric dipole have relatively weak electron scattering resulting in low signal-to-noise ratio, especially for HAADF images. In room-temperature bulk PbTiO<sub>3</sub>, the Ti B-site is offset by 16.2 pm and the oxygen by 47–49 pm relative to the Pb lattice in the direction opposite to the polarization<sup>35</sup>. Thus, we use the Ti to infer the offset of the Ti-'centred' oxygen octahedra for defining the electric dipole. Displacement vectors ( $P_{PD}$ ) corresponding to local offsets between the A- and B-site sublattices were calculated by first determining atomic positions by fitting each atom site in the image to four-parameter, spherical Gaussians using a trust-region algorithm in Matlab (see illustration in Extended Data Fig. 3c). Fits were performed simultaneously for five-atom clusters of the B-site-centred perovskite unit cell (Extended Data Fig. 3d). The displacement for each atom was then calculated as the difference between its atom position and the mean position (that is, centre of mass) of the surrounding four opposite-type cations (Extended Data Fig. 4c, d). Using a pseudocubic grid indexing (A-sites defining a grid) and defining the cross-section image to lie in the  $x$ - $z$  Cartesian plane ( $x = [100]_{pc}$ ,  $z = [001]_{pc}$ ), for an atom at grid position  $i, j$  this corresponds to mean neighbour positions (MNP) of:

$$\text{MNP}x_{ij} = (x_{i-0.5,j-0.5} + x_{i-0.5,j+0.5} + x_{i+0.5,j+0.5} + x_{i+0.5,j-0.5})/4$$

$$\text{MNP}z_{ij} = (z_{i-0.5,j-0.5} + z_{i-0.5,j+0.5} + z_{i+0.5,j+0.5} + z_{i+0.5,j-0.5})/4$$

Illustrative examples for both A-site and B-site regions are shown in Extended Data Fig. 4. To maintain a consistent displacement vector sign, we reverse the order for A-sites and B-sites such that:

$$P_{PD,ij \text{ A-site}} = (x_{ij} - \text{MNP}x_{ij})\hat{x} + (z_{ij} - \text{MNP}z_{ij})\hat{z}$$

$$P_{PD,ij \text{ B-site}} = (\text{MNP}x_{ij} - x_{ij})\hat{x} + (\text{MNP}z_{ij} - z_{ij})\hat{z}$$

(as illustrated in Extended Data Fig. 4d). This convention defines a displacement vector direction nominally parallel to the full electrical polarization of the bulk crystal structure. Noise in the displacement vector fields was reduced by weighted smoothing of the first derivative over a local range of 12 atoms (calculation details below).

A conservative estimation of the experimental error in the polar displacement maps for the data set in this work was performed by statistical analysis of the distribution of the polar displacements within the SrTiO<sub>3</sub> layers.  $P_{PD}$  for the Sr and Ti atoms within the SrTiO<sub>3</sub> layers (excluding the interfaces) has an  $\sigma_{RMS} = 4.88$  pm. This represents an upper bound on the experimental error, as it is predicated on all measured non-centrosymmetric distortions of SrTiO<sub>3</sub> being the result of noise. Any actual distortions, be they induced by the PbTiO<sub>3</sub> epitaxy or resulting from defects, are falsely included in this estimate.

The curl vector (or vorticity) of the displacement vector field  $P_{PD}$  in the TEM plane is the  $y$  component of  $\nabla \times P_{PD}$ , or  $\left(\frac{\partial P_{PD,z}}{\partial x} - \frac{\partial P_{PD,x}}{\partial z}\right)\hat{y}$  (again the TEM image is defined as the  $x$ - $z$  plane,  $y$  is the direction of the electron beam). The polarization gradient terms  $\frac{\partial P_{PD,z}}{\partial x}$  and  $\frac{\partial P_{PD,x}}{\partial z}$  were calculated by a weighted least-squares fit of a two-variable ( $x$  and  $z$ ) first-order polynomial to  $P_{PD,x}$  and  $P_{PD,z}$  at each atom site

with a fit area extending to 12 surrounding atoms. More succinctly, the local  $P_{\text{PD},x}$  and  $P_{\text{PD},z}$  values were fitted as a plane, and the derivatives are taken as the  $x$  and  $z$  slopes of the plane. A Gaussian function with a  $\sigma$  of the mean grid spacing was used as the weighting function, that is, the central atom was given a weight of 1 and decreased to a weight of  $\sim 0.1$  for the most distant of the 12 neighbour atoms considered. The polynomial fit was used as a smoothing function for the vector fields displayed in the figures, effectively smoothing the first derivative, and the gradients input into the curl equation. The curl vectors calculated and smoothed in this manner for the HAADF image in Fig. 2 are shown in Extended Data Fig. 9a for all atom positions. A comparison between the smoothed and raw polar displacements for one of the vortices is shown in Extended Data Fig. 9b and d, respectively.

**Phase-field simulations for superlattice structures.** In the phase-field modelling of the superlattice films, the evolution of the polarization was obtained by solving the time-dependent Ginzburg-Landau (TDGL) equations:

$$\frac{\partial P_i(\mathbf{r}, t)}{\partial t} = -L \frac{\delta F}{\delta P_i(\mathbf{r}, t)} \quad (i = 1, 2, 3) \quad (1)$$

where  $L$ ,  $\mathbf{r}$ , and  $t$  denote the kinetic coefficient, spatial position vectors, and time, respectively. The contributions to the total free energy  $F$  include the Landau bulk energy, elastic energy, electric (electrostatic) energy, and gradient energy, that is:

$$F = \int (f_{\text{Landau}} + f_{\text{elastic}} + f_{\text{electric}} + f_{\text{gradient}}) dV \quad (2)$$

The expression of each energy density can be found in the literature<sup>36–38</sup>. Owing to the inhomogeneity of elastic constants in the  $(\text{SrTiO}_3)_n/(\text{PbTiO}_3)_n$  superlattices, a spectral iterative perturbation method was employed to solve the mechanical equilibrium equation to obtain the stress field<sup>39</sup>. The pseudocubic lattice constants for  $\text{PbTiO}_3$  and  $\text{SrTiO}_3$  were taken as 3.9547 Å and 3.905 Å, respectively<sup>40</sup>, while the anisotropic in-plane lattice constants for the substrate  $\text{DyScO}_3$  is taken from literature to calculate the misfit strain<sup>33</sup>. Material constants for  $\text{PbTiO}_3$  and  $\text{SrTiO}_3$  used in the simulations are found in the literature and these include the Landau potentials, elastic constants, electrostrictive coefficients, background dielectric constants, and gradient energy coefficients<sup>37,40–44</sup>.

The 3D phase-field simulation of the  $(\text{SrTiO}_3)_{10}/(\text{PbTiO}_3)_{10}$  superlattice was done using discrete grids of  $(200\Delta x) \times (200\Delta y) \times (250\Delta z)$  with  $\Delta x = \Delta y = \Delta z = 0.4$  nm, where  $\Delta x$ ,  $\Delta y$ , and  $\Delta z$  are in real space. The thickness of the substrate, film, and air are  $(30\Delta z)$ ,  $(198\Delta z)$ , and  $(22\Delta z)$ , respectively. In the film, alternating  $n$  grids of  $\text{PbTiO}_3$  layers and  $n$  grids of  $\text{SrTiO}_3$  layers are incorporated to simulate the periodically layered superlattice, where  $n$  is the superlattice periodicity (for example,  $n = 10$  for the  $10 \times 10$  superlattice). Periodic boundary conditions are assumed in both the  $x$  and  $y$  directions, and a superposition method is used along the  $z$  direction<sup>37</sup>. Random noise is used as the initial setup to simulate the thermal fluctuation during the annealing process.

Using this approach, the 3D phase-field simulation was carried out for a  $(\text{SrTiO}_3)_{10}/(\text{PbTiO}_3)_{10}$  superlattice, and the results are shown in Fig. 4. The vortex-antivortex pairs extending along  $\text{DyScO}_3$   $[010]_{\text{pc}}$  are highlighted in Fig. 4 by red and blue colours having opposite vorticity. The long-range ordering of these vortex pairs along  $\text{DyScO}_3$   $[100]_{\text{pc}}$  in a  $10 \times 10$  superlattice can be clearly seen in Extended Data Fig. 5c; a subsection shows a vortex-antivortex pair in Extended Data Fig. 5d, which is also shown in Fig. 2d.

A comparison of phase-field models for  $n = 10$  and  $n = 50$   $\text{PbTiO}_3$  thickness highlights the difference between the vortex and flux-closure domain structure (Extended Data Fig. 7). In a true vortex state, polarization is continuously changing about a vortex core giving rise to an inhomogeneous polarization distribution. Such a polarization state is demonstrated using both HR-STEM displacement

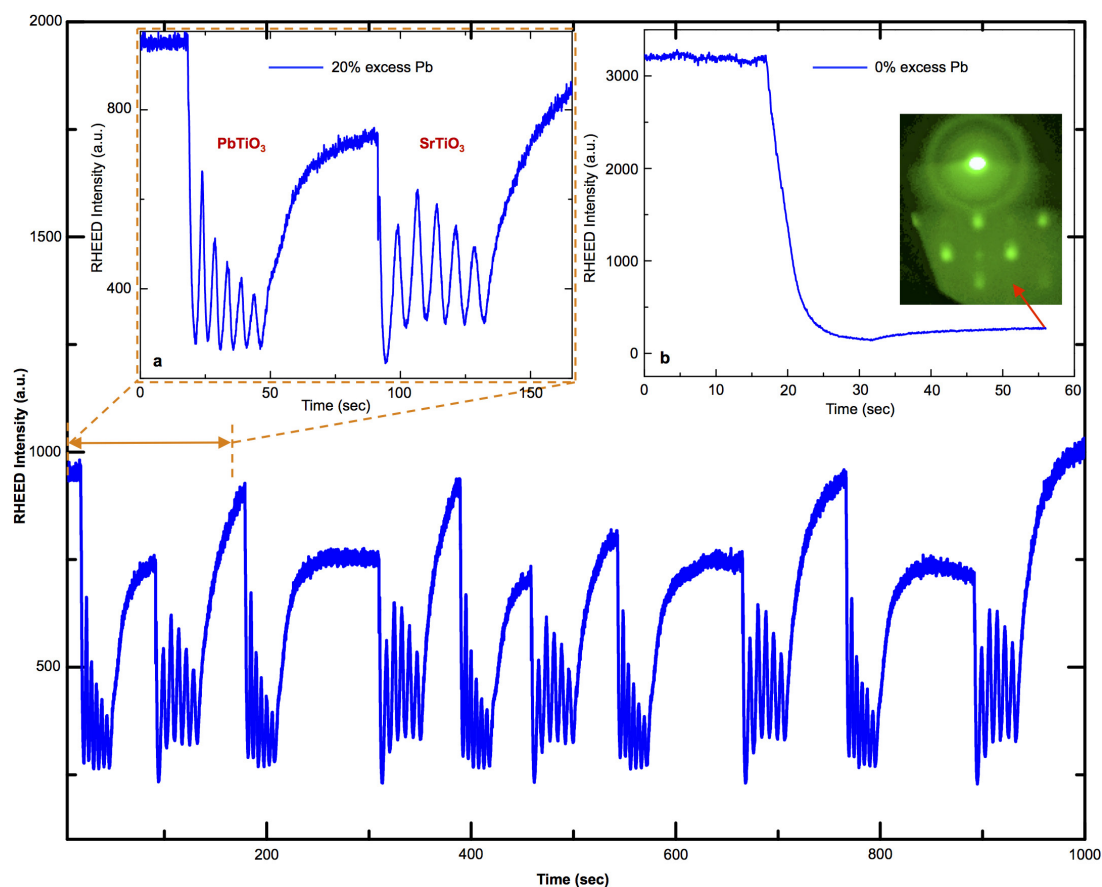
vector mapping (Extended Data Fig. 7a) and phase-field simulations (Extended Data Fig. 7b) for a  $10 \times 10$  superlattice. In a flux-closure domain (Extended Data Fig. 7d, e), on the other hand, regions of homogeneous polarization (with a net polarization) can be identified, which are separated by domain walls, giving rise to a polarization distribution similar to those previously observed in bulk materials<sup>18</sup>. In ref. 18, free-standing single crystal lamellae of  $\text{BaTiO}_3$  were probed using piezoresponse force microscopy, and the results show formation of flux-closure domains, at the length scale of micrometres, which are very similar to the domain pattern reported in ref. 19.

Phase-field calculations were further employed to complete a phase diagram of the evolution of vortex states in these superlattices (that is, total energy density as a function of superlattice periodicity; Extended Data Fig. 8). At lower periodicities ( $< 10 \times 10$ ), calculations predict the existence of ferroelastic domains, shown as  $a_1/a_2$  by solid squares.  $a_1/a_2$ -type domain structure have domains which are polarized along the in-plane directions (the directions along which mechanical constraints are applied, in this case along  $[100]_{\text{pc}}$  and  $[010]_{\text{pc}}$ ), and are separated by  $90^\circ$  domain walls, making them ferroelastic in nature<sup>45</sup>. After reaching a critical periodicity ( $\sim 10 \times 10$ ), total energy calculations show that the vortex structure becomes more favourable as compared to the  $a_1/a_2$ . Upon transitioning to larger superlattice periods ( $> 27 \times 27$ ), the polarization state is stabilized by incorporating flux-closure-type domains, where regions of homogeneous polarization exist, and are separated by domain walls.

**Sample size.** No statistical methods were used to predetermine sample size.

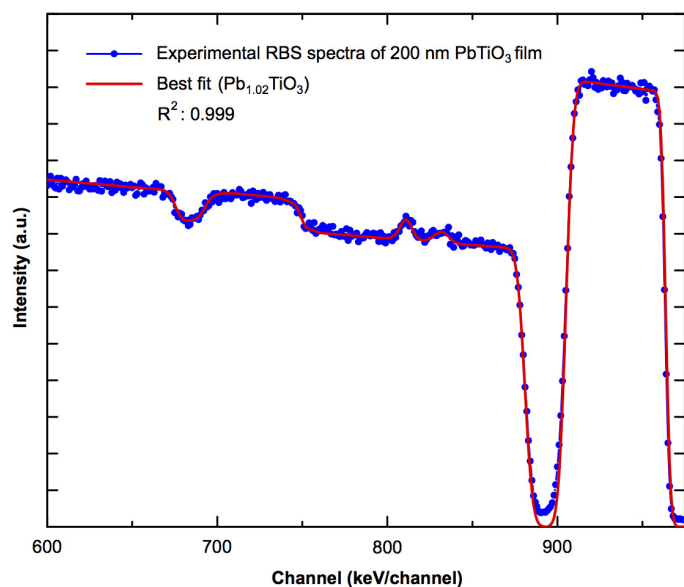
- Kuroiwa, Y., Aoyagi, S. & Sawada, A. Evidence of Pb-O covalency in tetragonal  $\text{PbTiO}_3$ . *Phys. Rev. Lett.* **87**, 217601 (2001).
- Jones, C. W., Battle, P. D., Lightfoot, P. & Harrison, T. A. The structure of  $\text{SrRuO}_3$  by time-of-flight neutron powder diffraction. *Acta Crystallogr. C* **45**, 365–367 (1989).
- Chen, Z. H., Damodaran, A. R., Xu, R., Lee, S. & Martin, L. W. Effect of “symmetry mismatch” on the domain structure of rhombohedral  $\text{BiFeO}_3$  thin films. *Appl. Phys. Lett.* **104**, 182908 (2014).
- Ophus, C., Nelson, C. T. & Ciston, J. Correcting nonlinear drift distortion of scanning probe microscopy from image pairs with orthogonal scan directions. *Ultramicroscopy* **162**, 1–9 (2016).
- Glazer, A. M. & Mabud, S. A. Powder profile refinement of lead zirconium titanate at several temperatures. II.  $\text{PbTiO}_3$ . *Acta Crystallogr. B* **34**, 1065–1070 (1978).
- Chen, L.-Q. Phase-field model of phase transitions/domain structures in ferroelectric thin films: a review. *J. Am. Ceram. Soc.* **91**, 1835–1844 (2008).
- Li, Y. L., Hu, S. Y., Liu, Z. K. & Chen, L.-Q. Effect of substrate constraint on the stability and evolution of ferroelectric domain structures in thin films. *Acta Mater.* **50**, 395–411 (2002).
- Li, Y. L., Hu, S. Y., Liu, Z. K. & Chen, L.-Q. Effect of electrical boundary conditions on ferroelectric domain structures in thin films. *Appl. Phys. Lett.* **81**, 427–429 (2002).
- Wang, J. J., Ma, X. Q., Li, Q., Britson, J. & Chen, L.-Q. Phase transitions and domain structures of ferroelectric nanoparticles: phase field model incorporating strong elastic and dielectric inhomogeneity. *Acta Mater.* **61**, 7591–7603 (2013).
- Haun, M. J., Furman, E., Jiang, S. J., McKinstry, H. A. & Cross, L. E. Thermodynamic theory of  $\text{PbTiO}_3$ . *J. Appl. Phys.* **62**, 3331–3338 (1987).
- Sheng, G. et al. A modified Landau-Devonshire thermodynamic potential for strontium titanate. *Appl. Phys. Lett.* **96**, 232902 (2010).
- Tagantsev, A. K. Landau expansion for ferroelectrics: which variable to use? *Ferroelectrics* **375**, 19–27 (2008).
- Tagantsev, A. K. The role of background dielectric susceptibility in uniaxial ferroelectrics. *Ferroelectrics* **69**, 321–323 (1986).
- Zheng, Y. & Woo, C. H. Giant piezoelectric resistance in ferroelectric tunnel junctions. *Nanotechnology* **20**, 075401 (2009).
- Pompe, W., Gong, X., Suo, Z. & Speck, J. S. Elastic energy release due to domain formation in the strained epitaxy of ferroelectric and ferroelastic films. *J. Appl. Phys.* **74**, 6012–6019 (1993).



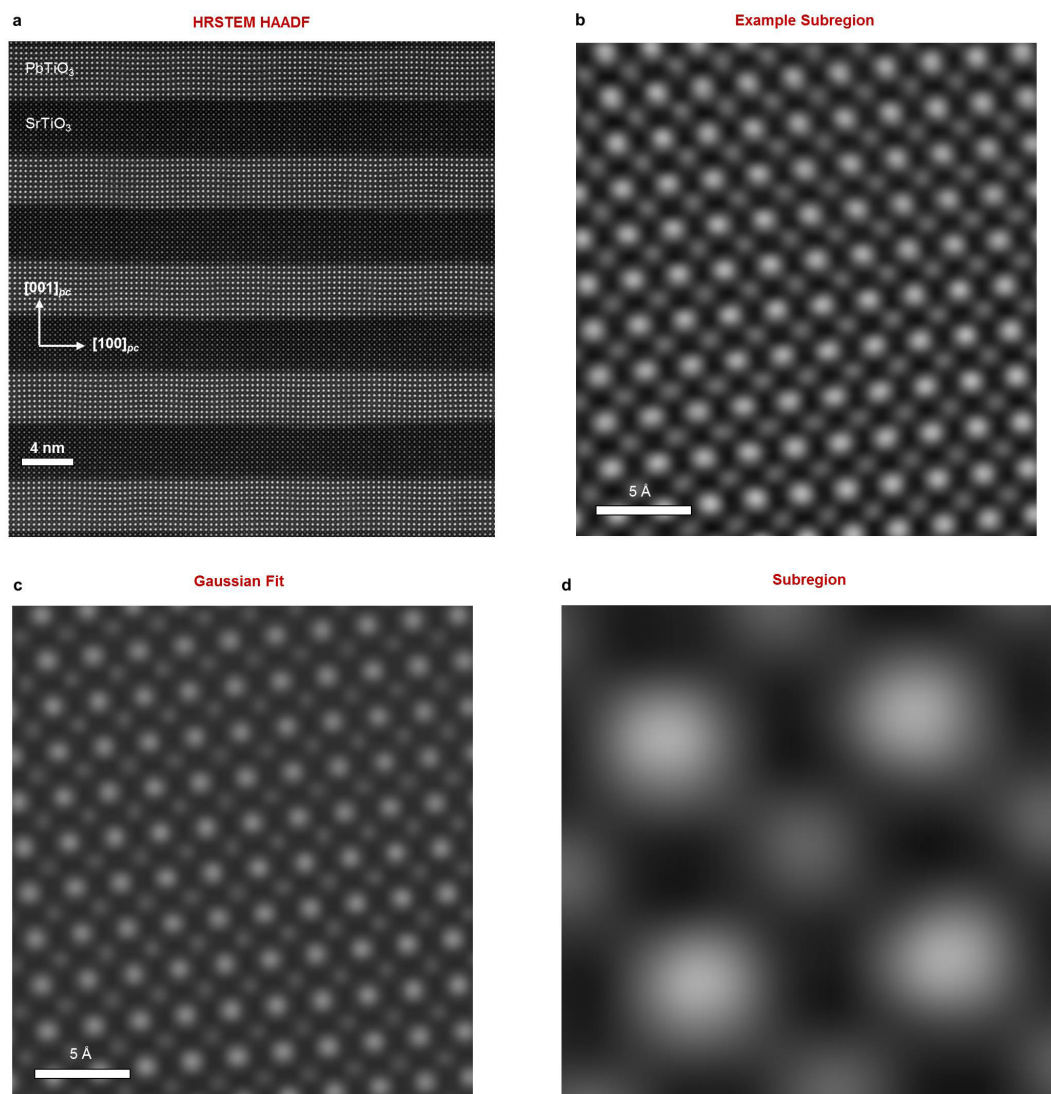


**Extended Data Figure 1 | RHEED-controlled growth of superlattices.** RHEED oscillations for five periods of a  $(\text{SrTiO}_3)_6/(\text{PbTiO}_3)_6$  superlattice grown using a 20% excess Pb target are shown. The oscillations were present throughout the growth of the 100-nm-thick superlattice. **a**, RHEED oscillations for six unit cells of  $\text{PbTiO}_3$ , followed by oscillations

for six unit cells of  $\text{SrTiO}_3$ , that is, one  $6 \times 6$  superlattice period. **b**, RHEED intensity variation for growth of  $\text{PbTiO}_3$  from a stoichiometric  $\text{PbTiO}_3$  or 0% Pb excess target. Inset, RHEED diffraction pattern obtained at the end of growth, highlighting the rapid transition to 3D growth mode when a stoichiometric target is used (Methods).



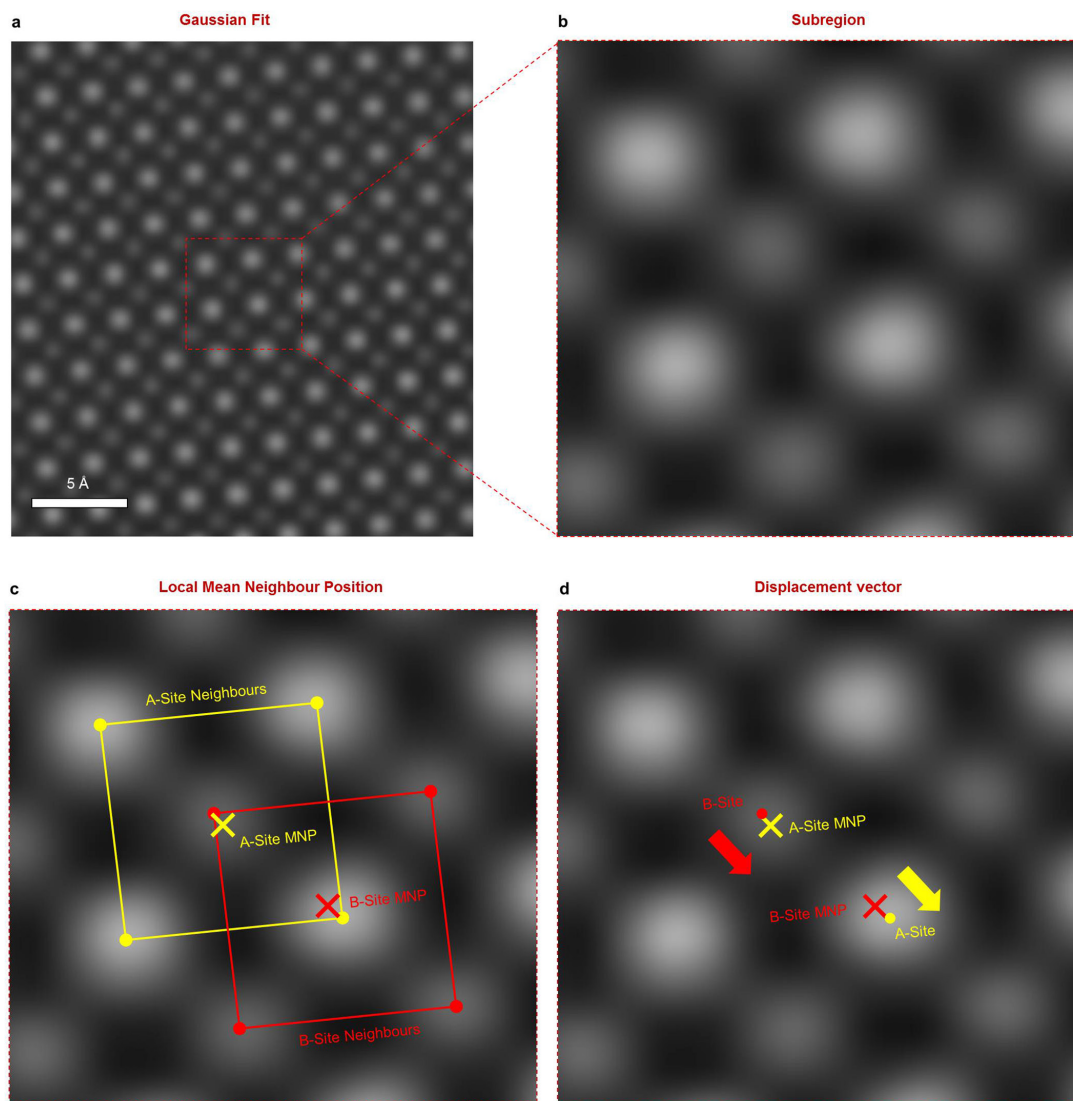
**Extended Data Figure 2 | Rutherford backscattering spectrometry (RBS) characterization.** An RBS spectrum of a 200-nm  $\text{PbTiO}_3$  film grown on a single-crystalline  $\text{GdScO}_3(001)_{\text{pc}}$  substrate is shown, along with the simulated spectrum (indicated by 'Best fit' in the figure) for the same structure, revealing an effective Pb:Ti ratio of 1.02:1, which is very close to ideal stoichiometry.



**Extended Data Figure 3 | Fitting atomic positions in the HR-STEM images.** **a**, An HR-STEM HAADF image of a  $(\text{SrTiO}_3)_{10}/(\text{PbTiO}_3)_{10}$  superlattice. Polar displacement vectors ( $P_{\text{PD}}$ ) are calculated by first finding atom positions by fits of 2D Gaussians to the image intensity.

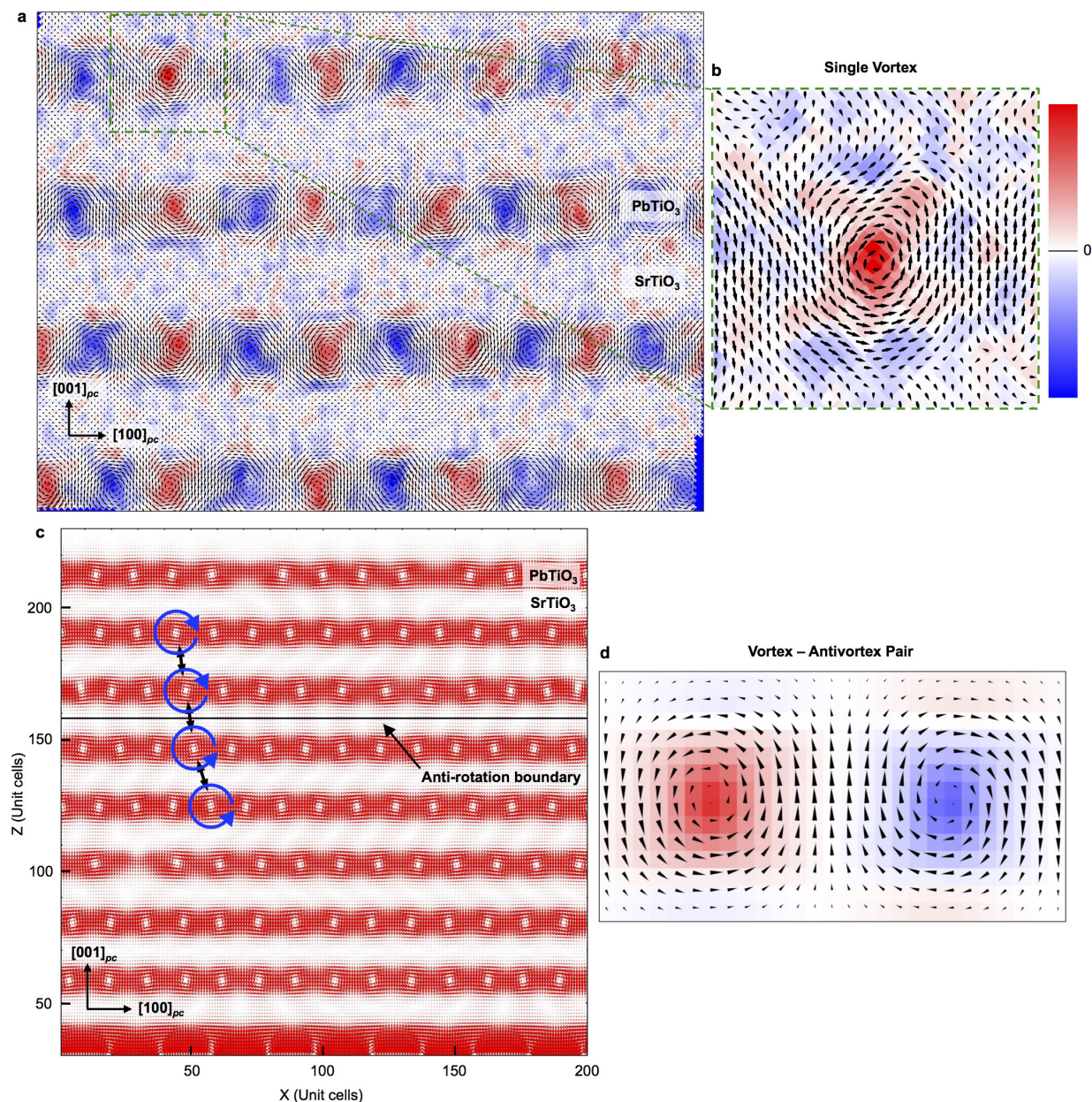
**b**, An example subregion from an HR-STEM image (unrotated). **c**, The corresponding Gaussian fit of **b**. **d**, An example of the B-site-centred five-atom clusters which are fitted simultaneously.





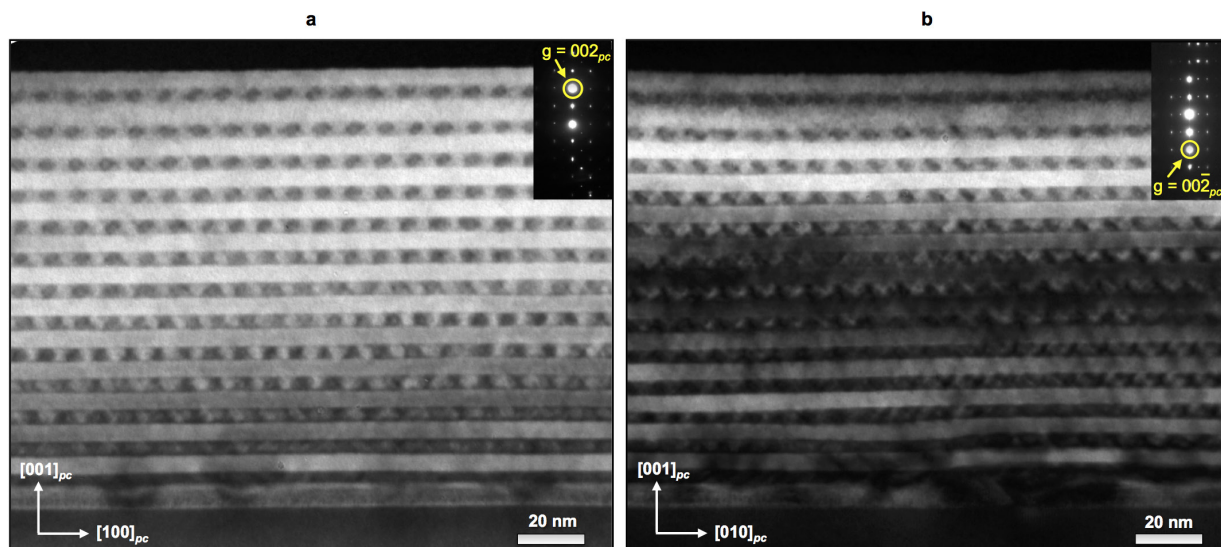
**Extended Data Figure 4 | Extracting polar displacement vectors ( $P_{PD}$ ) from HR-STEM data.** **a**, The Gaussian fit of a region of an HAADF STEM image. **b**, A magnified view of the boxed subregion in **a**, showing two overlapping unit cells, namely, the A-site-centred perovskite unit cell, and the B-site-centred perovskite unit cell. **c**, The mean position of each atom's

four nearest cation neighbours (MNP, mean neighbour position) is used to determine a relative sublattice offset. **d**, A polar displacement vector (yellow arrow for A-site and red arrow for B-site), taken as the difference between each atom (filled circle) and the MNP ('x'), is given opposite sign for A- and B-sites to maintain a consistent direction.



**Extended Data Figure 5 | Polar displacement vector map and phase-field simulation of a  $(\text{SrTiO}_3)_{10}/(\text{PbTiO}_3)_{10}$  superlattice.** **a**, The polar displacement vector map (black arrows) corresponding to the  $(\text{SrTiO}_3)_{10}/(\text{PbTiO}_3)_{10}$  superlattice in Fig. 2a is shown overlaid on the colorized  $(\nabla \times \mathbf{P}_{\text{PD}})|_{[010]}$ . **b**, A magnified view of the highlighted polar vortex shows the smooth rotation increasing towards a vortex core. The extremes of the colour scale of the curl vector are  $-16.75 \text{ pm nm}^{-1}$  and  $16.75 \text{ pm nm}^{-1}$ , with blue and red colours indicating negative and positive values of curl

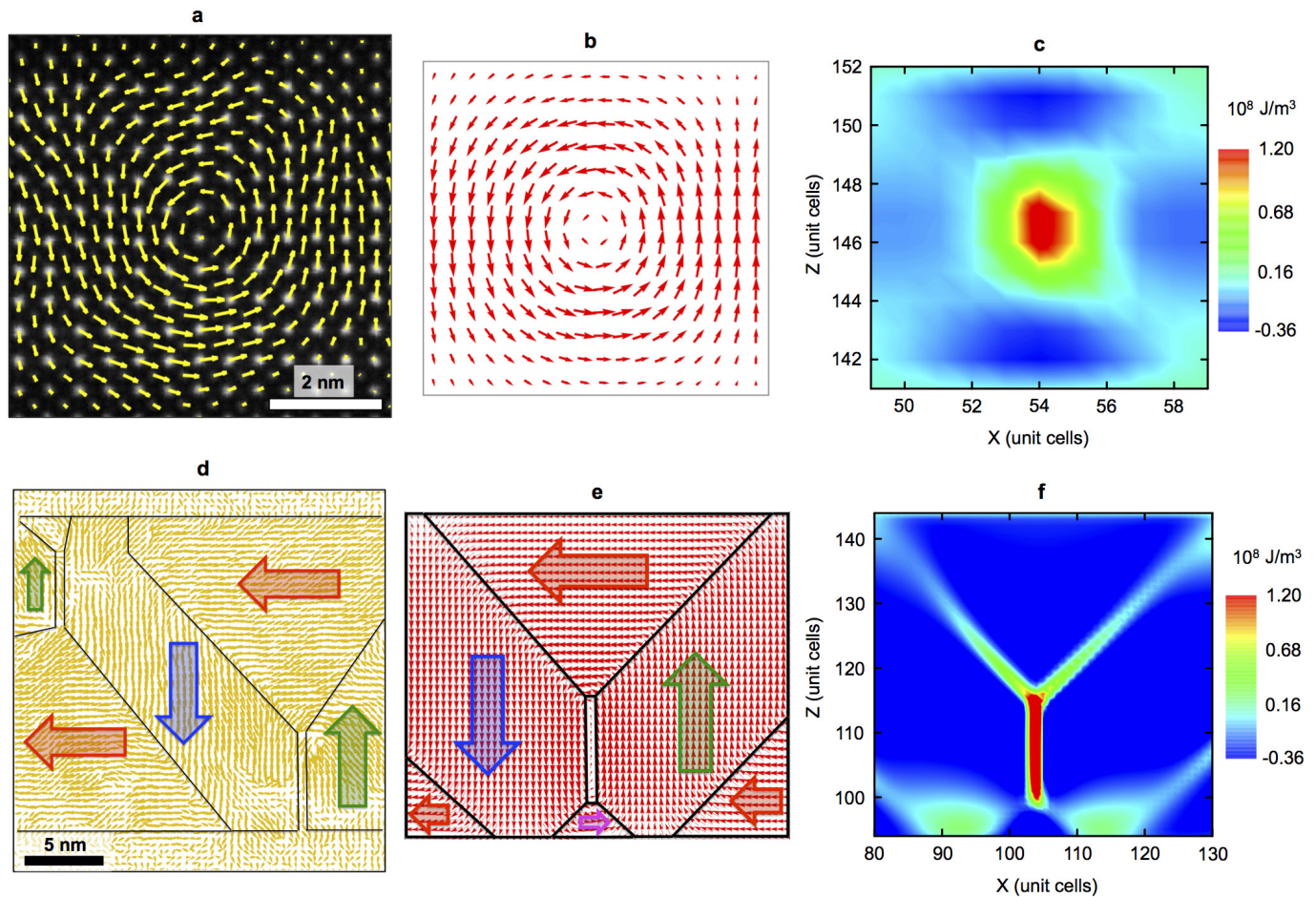
(units for polar displacement are in pm). **c**, A cross-section of the phase-field simulation for the  $(\text{SrTiO}_3)_{10}/(\text{PbTiO}_3)_{10}$  superlattice showing the polarization vectors (red arrows) with interlayer alignment. This sample exhibits preference for alignment of matched rotations (indicated by blue arrows), but a phase shift of anti-alignment is also found (indicated by 'Anti-rotation boundary' in the figure). Such defects are found occasionally. **d**, A subregion of the simulation, highlighting a vortex-antivortex pair with the curl of the polarization overlaid.



**Extended Data Figure 6 | Vortex-antivortex arrays along both pseudo-cubic in-plane directions.** **a**, Cross-sectional DF-TEM of a  $(\text{SrTiO}_3)_{10}/(\text{PbTiO}_3)_{10}$  superlattice taken along the  $[010]_{pc}$  zone axis (inset shows the SAED pattern and the  $g$ -vector for image formation). **b**, Cross-sectional

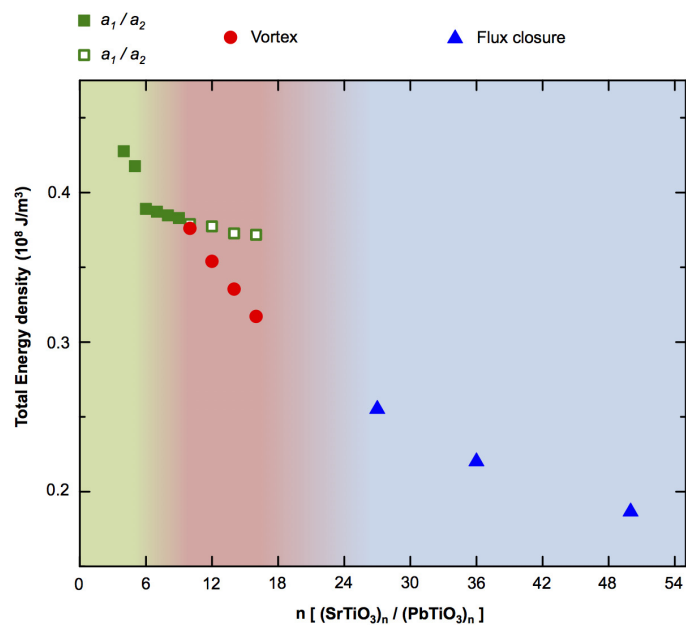
DF-TEM of the same  $(\text{SrTiO}_3)_{10}/(\text{PbTiO}_3)_{10}$  superlattice sample taken along the orthogonal  $[100]_{pc}$  zone axis (inset as **a**). In both cases, clear vortex-antivortex structures are observed.



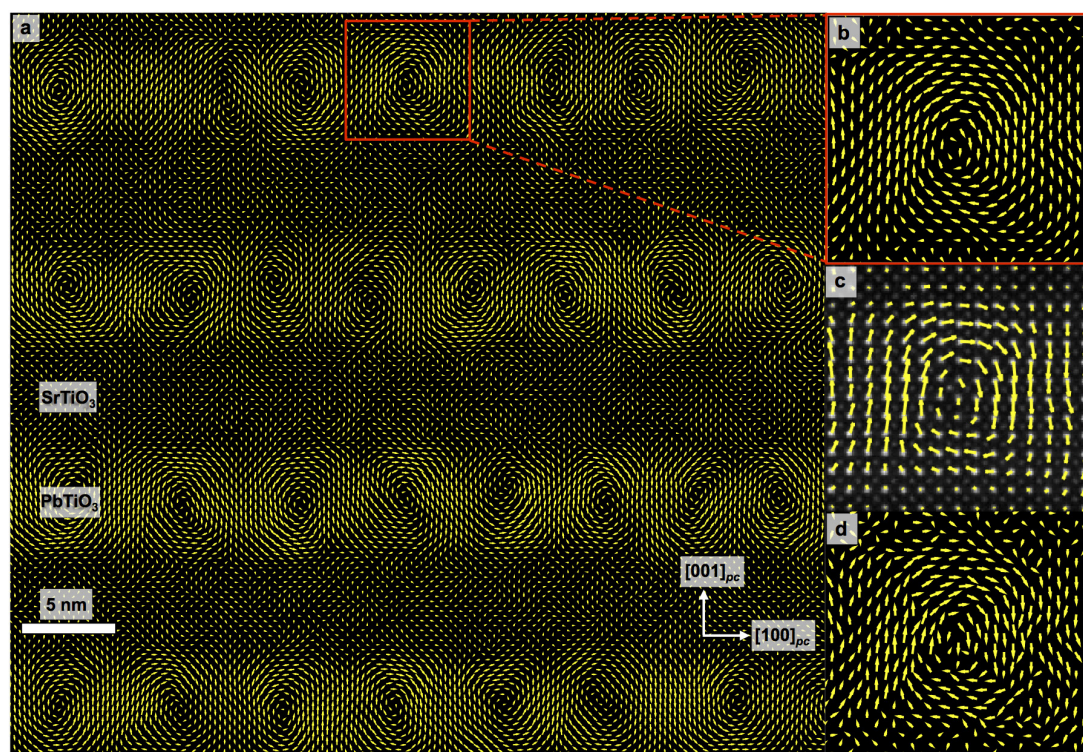


**Extended Data Figure 7 | Differentiation of vortex structure from a flux-closure domain structure using phase-field simulations and HR-STEM polar displacement mapping.** **a**, HR-STEM polar displacement vector map of a single vortex in a  $(\text{SrTiO}_3)_{10}/(\text{PbTiO}_3)_{10}$  superlattice. **b**, Phase-field-calculated polarization vector map of a single vortex in a  $(\text{SrTiO}_3)_{10}/(\text{PbTiO}_3)_{10}$  superlattice. **c**, Total energy density variation within a single vortex in a  $(\text{SrTiO}_3)_{10}/(\text{PbTiO}_3)_{10}$  superlattice calculated from phase-field simulations. **d**, HR-STEM polar displacement

vector map of a 20-nm  $\text{PbTiO}_3$  layer in a heterostructure having the same layer stacking as in ref. 19. **e**, Phase-field-calculated polarization vector map of a flux-closure domain in a 20-nm  $\text{PbTiO}_3$  heterostructure (red, blue, green and pink arrows in **d** and **e** indicate the direction of net polarization within their respective domains). **f**, Total energy density variation within a flux-closure domain calculated using phase-field simulations.



**Extended Data Figure 8 | Phase evolution of vortex structure.** The phase-field-calculated total energy for various phases competing with a vortex phase is depicted. See Methods for details.



**Extended Data Figure 9 | HR-STEM displacement vector map of a  $(\text{SrTiO}_3)_{10}/(\text{PbTiO}_3)_{10}$  superlattice.** **a**, Polar displacement vector map of a  $(\text{SrTiO}_3)_{10}/(\text{PbTiO}_3)_{10}$  superlattice obtained from the same HR-STEM data set as Fig. 2a, showing the polar displacement vectors at all atom

positions. **b**, A magnified image of a single vortex. **c**, Polar displacement vector map of the same region as **b**, but plotting only Pb (A-site) centred displacements (like Fig. 2a). **d**, Polar displacement vector map of a single vortex showing the raw displacement vectors.



# Macromolecular diffractive imaging using imperfect crystals

Kartik Ayyer<sup>1</sup>, Oleksandr M. Yefanov<sup>1</sup>, Dominik Oberthür<sup>2</sup>, Shatabdi Roy-Chowdhury<sup>3,4</sup>, Lorenzo Galli<sup>1,2</sup>, Valerio Mariani<sup>1</sup>, Shibom Basu<sup>3,4</sup>, Jesse Coe<sup>3,4</sup>, Chelsie E. Conrad<sup>3,4</sup>, Raimund Fromme<sup>3,4</sup>, Alexander Schaffer<sup>3,4</sup>, Katerina Dörner<sup>1,3</sup>, Daniel James<sup>4,5</sup>, Christopher Kupitz<sup>3,6</sup>, Markus Metz<sup>2</sup>, Garrett Nelson<sup>4,5</sup>, Paulraj Lourdu Xavier<sup>1,2</sup>, Kenneth R. Beyerlein<sup>1</sup>, Marius Schmidt<sup>6</sup>, Iosifina Sarrou<sup>7</sup>, John C. H. Spence<sup>4,5</sup>, Uwe Weierstall<sup>4,5</sup>, Thomas A. White<sup>1</sup>, Jay-How Yang<sup>3,4</sup>, Yun Zhao<sup>4,5</sup>, Mengning Liang<sup>8</sup>, Andrew Aquila<sup>8</sup>, Mark S. Hunter<sup>8</sup>, Joseph S. Robinson<sup>8</sup>, Jason E. Koglin<sup>8</sup>, Sébastien Boutet<sup>8</sup>, Petra Fromme<sup>3,4</sup>, Anton Barty<sup>1</sup> & Henry N. Chapman<sup>1,2,9</sup>

**The three-dimensional structures of macromolecules and their complexes are mainly elucidated by X-ray protein crystallography. A major limitation of this method is access to high-quality crystals, which is necessary to ensure X-ray diffraction extends to sufficiently large scattering angles and hence yields information of sufficiently high resolution with which to solve the crystal structure. The observation that crystals with reduced unit-cell volumes and tighter macromolecular packing often produce higher-resolution Bragg peaks<sup>1,2</sup> suggests that crystallographic resolution for some macromolecules may be limited not by their heterogeneity, but by a deviation of strict positional ordering of the crystalline lattice. Such displacements of molecules from the ideal lattice give rise to a continuous diffraction pattern that is equal to the incoherent sum of diffraction from rigid individual molecular complexes aligned along several discrete crystallographic orientations and that, consequently, contains more information than Bragg peaks alone<sup>3</sup>. Although such continuous diffraction patterns have long been observed—and are of interest as a source of information about the dynamics of proteins<sup>4</sup>—they have not been used for structure determination. Here we show for crystals of the integral membrane protein complex photosystem II that lattice disorder increases the information content and the resolution of the diffraction pattern well beyond the 4.5-ångström limit of measurable Bragg peaks, which allows us to phase<sup>5</sup> the pattern directly. Using the molecular envelope conventionally determined at 4.5 ångströms as a constraint, we obtain a static image of the photosystem II dimer at a resolution of 3.5 ångströms. This result shows that continuous diffraction can be used to overcome what have long been supposed to be the resolution limits of macromolecular crystallography, using a method that exploits commonly encountered imperfect crystals and enables model-free phasing<sup>6,7</sup>.**

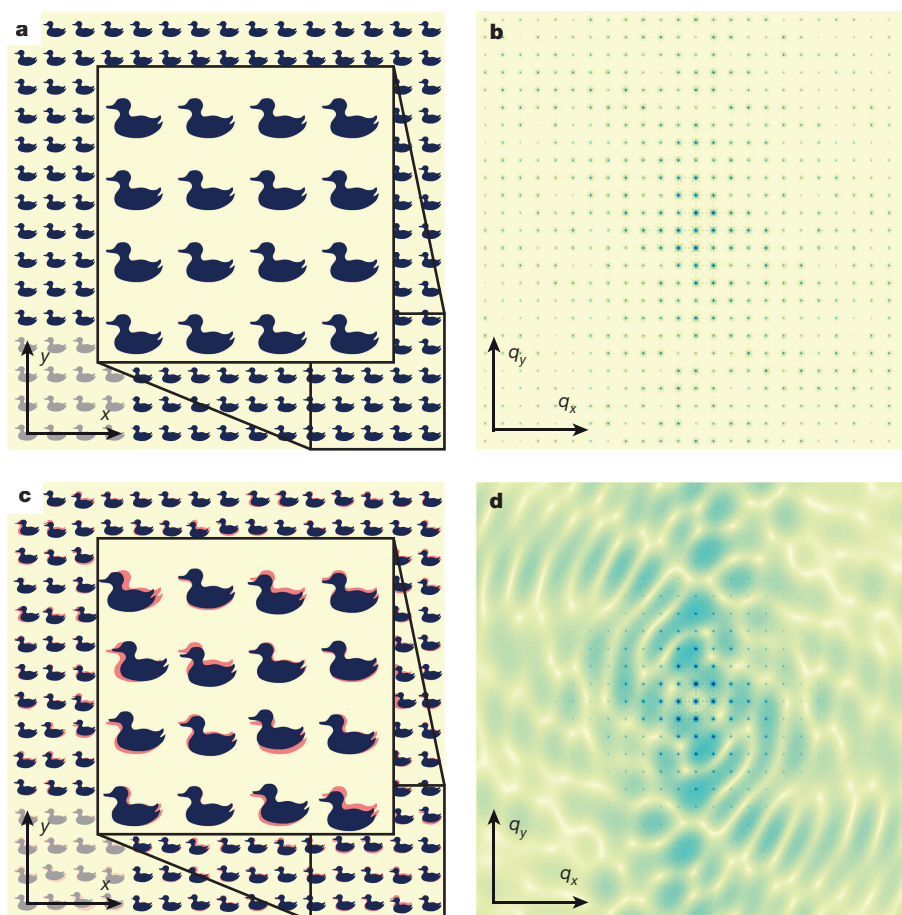
High-resolution Bragg diffraction from a crystal requires a high degree of regularity, which is often not obtained in practice. If one of the structural units that make up a crystal (such as a molecule) is displaced from the ideal lattice by an amount  $s$ , then the phase of the diffracted wave from this unit is changed by  $2\pi s/d$  at a scattering angle  $2\theta$  corresponding to the resolution  $d$ , leading to destructive interference at that resolution when  $s = d/2$ . If all units are displaced randomly with a root-mean-square (r.m.s.) displacement  $\sigma$  along a given coordinate, then Bragg intensities will accordingly diminish at a resolution of  $d = 2\pi\sigma$  according to the well known Debye–Waller factor  $\exp(-4\pi^2 q^2 \sigma^2)$ , where  $q = 1/d = 2\sin(\theta/\lambda)$  for an X-ray wavelength  $\lambda$

(ref. 8). The disorder length  $\sigma$  is easily estimated from a so-called Wilson plot<sup>9</sup> of the diffraction data. Less widely appreciated is the fact that although the Bragg peaks diminish exponentially with the square of scattering angle, an incoherent sum of the diffraction of individual structural units arises contrariwise<sup>8,10,11</sup> (see Fig. 1 and Methods section ‘The effect of rigid body disorder on Bragg and continuous diffraction’) to conserve the energy lost from the Bragg peaks. This incoherent sum is equal to the Fraunhofer diffraction pattern of the structural unit multiplied by the number of units, without any modulation by Bragg peaks.

Continuously modulated diffraction intensities have frequently been observed in diffraction patterns of protein crystals<sup>10–15</sup>, and have even been ascribed to the disorder in the arrangement of rigid structural units as large as single macromolecules<sup>10–13</sup>. Although the continuous diffraction was seen to follow the point-group symmetry of the crystal<sup>14</sup>, previous studies did not consider the use of these patterns for structure determination. Protein crystals usually contain several molecules or molecular assemblies per unit cell as well as a solvent. The packing of molecules is mediated by non-covalent contacts between them (H bonds as well as electrostatic and hydrophobic interactions). It is therefore reasonable to propose a model in which the macromolecule is the rigid structural unit. In this case, with uncorrelated molecular displacements, the continuous diffraction pattern will consist of the incoherent sum of the continuous diffraction patterns of the macromolecule aligned in each of its various orientations within the crystal (given by equation (5) in Methods).

Accessing the ‘single molecule’ continuous diffraction overcomes the phase problem of crystallography, by increasing the number of observable coefficients to exceed the number required to describe the molecule, independent of resolution<sup>16,17</sup> (see Methods section ‘Constraint ratio’). Bragg intensities can be equivalently represented in real space, by Fourier transformation, as the three-dimensional (3D) pair-correlation function (or Patterson function<sup>9</sup>) of the crystal. This correlation function cannot always distinguish between object pairs existing only in one unit cell or straddling neighbours, but such an ambiguity is removed if the object is non-repeating. The highly sampled single-molecule diffraction data of an imperfect crystal (averaged over the orientational symmetries of the point group) could be used to obtain a real-space image of the macromolecule at a resolution beyond the resolution of the Bragg peaks, using iterative phasing algorithms<sup>18–22</sup> of coherent diffractive imaging and aligned molecule diffraction<sup>16</sup>.

<sup>1</sup>Center for Free-Electron Laser Science, DESY, 22607 Hamburg, Germany. <sup>2</sup>Department of Physics, University of Hamburg, 22761 Hamburg, Germany. <sup>3</sup>School of Molecular Sciences, Arizona State University, Tempe, Arizona 85287, USA. <sup>4</sup>Center for Applied Structural Discovery, Biodesign Institute, Arizona State University, Tempe, Arizona 85287, USA. <sup>5</sup>Department of Physics, Arizona State University, Tempe, Arizona 85287, USA. <sup>6</sup>Physics Department, University of Wisconsin, Milwaukee, Wisconsin 53211, USA. <sup>7</sup>Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology, Hellas, GR-70013 Crete, Greece. <sup>8</sup>Linac Coherent Light Source, Stanford Linear Accelerator Center (SLAC), National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California 94025, USA. <sup>9</sup>Centre for Ultrafast Imaging, 22607 Hamburg, Germany.



**Figure 1 | Lattice disorder reveals the continuous molecular transform.** **a, b**, Diffraction from an arrangement of objects in a perfect lattice (**a**) results in regularly spaced Bragg peaks (**b**) as a result of constructive interference. **c, d**, Translational disorder of the objects at a length scale  $\sigma$  (**c**) disrupts Bragg interference beyond a reciprocal-lattice resolution

length of  $d = 1/q = 2\pi\sigma$  (**d**). The loss of correlation gives rise to the incoherent sum of the Fraunhofer diffraction patterns of individual objects, which increases with  $q$  in balance with the diminishing Bragg intensities. Here,  $\sigma$  was chosen to be 4% of the lattice spacing, leading to reduction of Bragg peaks at the fourth order.

Measurements from crystals of the large integral membrane protein complex photosystem II (PSII) support the notion that the inherent resolution of the single asymmetric unit in the crystal can be better than that of the entire crystal. Recent X-ray free-electron laser (XFEL) experiments yielded Bragg diffraction from crystals with volumes  $1\mu\text{m}^3$  to a resolution of 5 Å (ref. 23). The protein-cofactor complex consists of a dimer of approximately 700 kDa. There are only four crystal contacts per dimer, and 64% solvent by volume in the crystal. If ascribed to disorder of the lattice alone, the Bragg limit of 5 Å could be caused by a r.m.s. displacement of molecules by 0.8 Å in each dimension (1.4-Å r.m.s. 3D displacement; see Methods section ‘The effect of rigid body disorder on Bragg and continuous diffraction’). Bragg diffraction from large PSII crystals ( $1\text{mm}^3$  in volume) extending to 1.9 Å was recently reported<sup>2</sup>, achieved by dehydrating the crystal in the presence of small amphiphiles, leading to an increase in crystal contacts and a tighter crystal packing. We set out to determine whether the limited crystallographic resolution of 5 Å is caused by random displacements of rigid-body structural units.

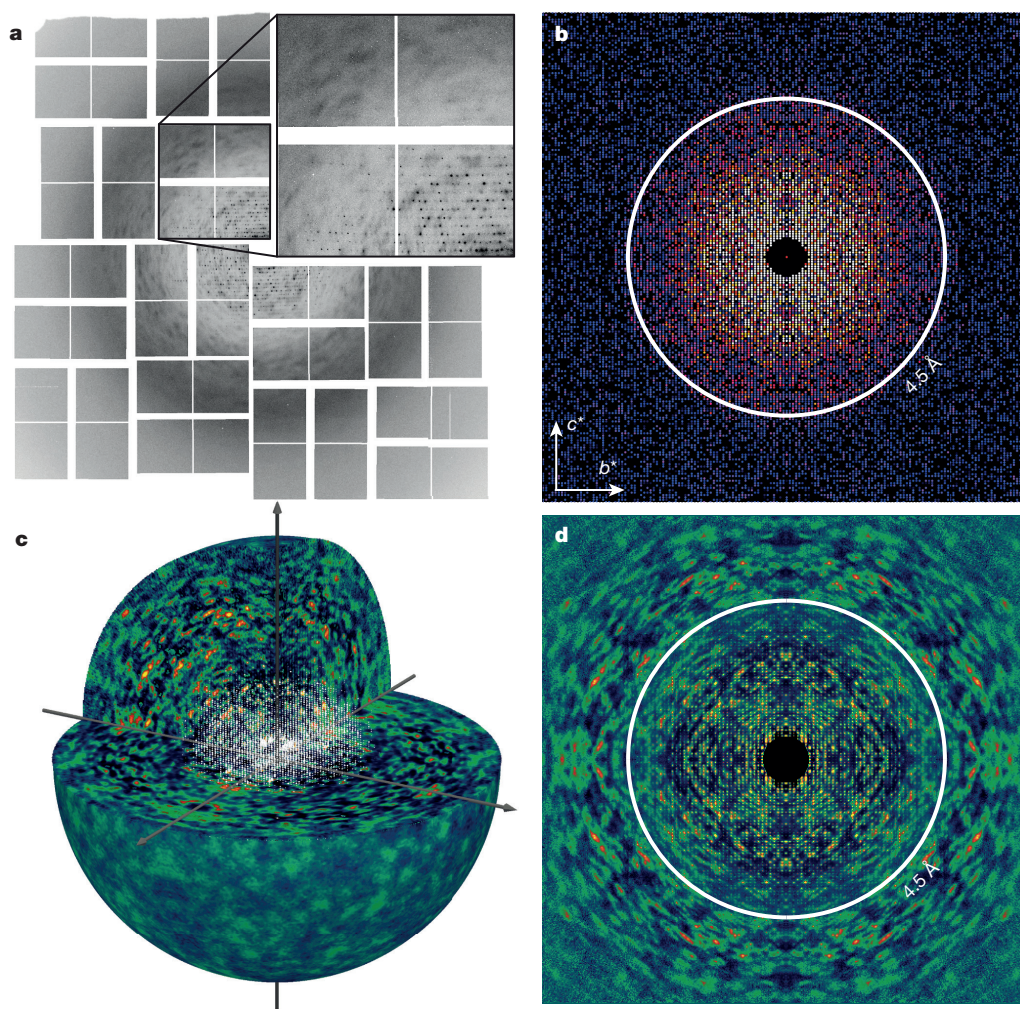
Still snapshot diffraction patterns, such as those shown in Fig. 2a, were recorded from room-temperature microcrystals of PSII in random orientations, as per the method of serial femtosecond crystallography<sup>23–25</sup> (see Methods section ‘Crystallography’). The orientation of each individual crystal was determined from its diffraction pattern by indexing the Bragg peaks<sup>26</sup>, allowing it to be inserted into a 3D array to build up the full intensity distribution<sup>27</sup> in the crystal coordinate frame (Fig. 2c, d). The accumulation of continuously modulated features, or

‘speckles’, in the diffraction volume implies that they are associated with the crystal rather than other scattering sources. Further, consistent with the hypothesis of rigid-body random displacements, the features are absent at the lowest resolutions and become dominant at scattering angles beyond the highest-angle Bragg peaks. The size of the structural unit can be estimated in an unbiased fashion from the autocorrelation of the diffracting object (the 3D pair-distribution function), which is obtained by a Fourier transform of the continuous diffraction intensities<sup>22</sup>. We found it to be in agreement with the size of a PSII dimer, as shown in Fig. 3.

One can consider several ways to obtain a high-resolution structure from the data shown in Fig. 2. We found that a particularly robust method was to treat the Bragg and continuous diffraction as two distinct sources of data of the same structure, the first arising from a coherent interference of molecules in the unit cell, and the second due to the incoherent addition of aligned single-molecule diffraction (see Methods section ‘Two data sources from one experiment’). Starting with a known search model for PSII, we first produced an electron density map of the PSII dimer by molecular replacement phasing to a resolution of 4.5 Å, limited by the angular extent of the Bragg peaks. This map was then used to generate a low-resolution binary mask of the smoothed molecular envelope of a single PSII dimer, shown in Fig. 3d. This mask forms the real-space constraint<sup>4</sup> used to generate a 3D image of the electron density of the PSII dimer by iterative phasing of the continuous diffraction.

The iterative phasing of the continuous diffraction, covering a resolution range of 4.5–3.3 Å, was carried out using the difference-map





**Figure 2 | Molecular coherent diffraction.** **a**, An XFEL snapshot ‘still’ diffraction pattern of a PSII microcrystal shows a weak speckle structure beyond the extent of Bragg peaks, which is enhanced in this figure by limiting the displayed pixel values. **b**, Structure factors obtained from Bragg-peak counts from 25,585 still patterns, displayed as a precession-style pattern of the [001] zone axis. **c**, A rendering of the

entire 3D diffraction volume assembled from the 2,848 strongest patterns. **d**, A central section of the diffraction volume in **c** normal to the [100] axis. Speckles are clearly observed beyond the 4.5-Å extent of Bragg diffraction (indicated by the white circles in **b** and **d**) to the edge of the detector.

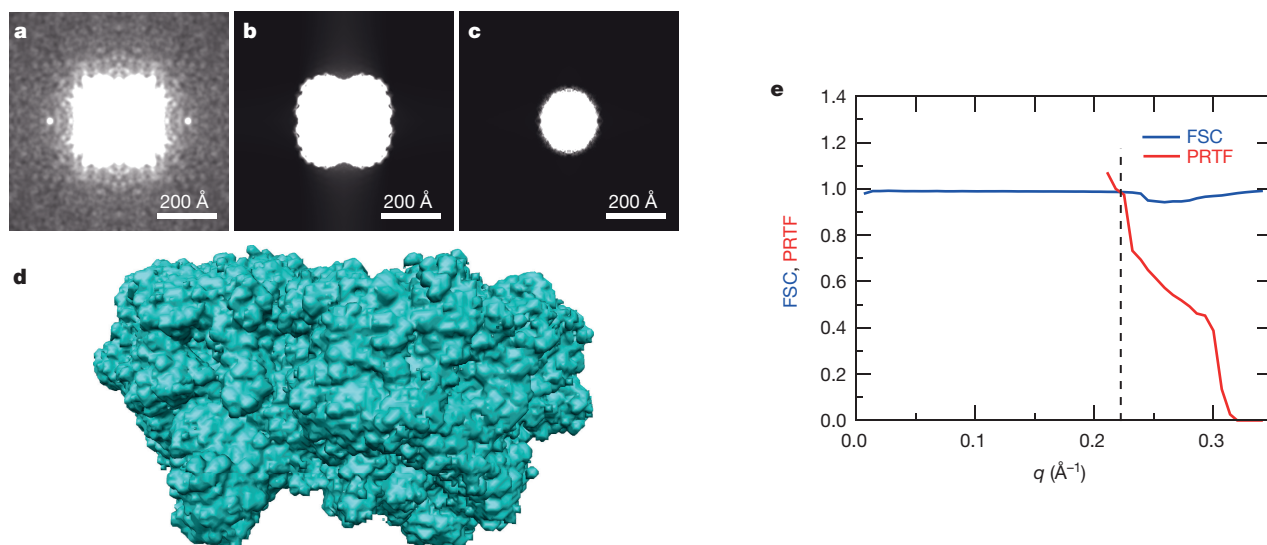
algorithm<sup>16,19</sup> (Methods section ‘Iterative phasing’). As with other iterative phasing algorithms, an initial guess of the electron density obtained from random phases is constrained on each iteration to be consistent with the measured diffraction and to fit within a certain finite (but not necessarily precisely known) real-space extent called the ‘support’—the smoothed envelope generated from the 4.5-Å resolution electron density map in this case.

After convergence of the iterate, the Fourier amplitudes and phases over the entire resolution range were combined to synthesize a 3.3-Å resolution structure. Averaging solutions obtained over multiple random starts produced a self-consistent electron density to a resolution of 3.5 Å, as determined by the Fourier shell correlation and the phase retrieval transfer function<sup>28</sup> (Fig. 3e). Finally, pseudo-crystallographic refinement<sup>29</sup> of this electron density was carried out, following a similar procedure as for single-particle cryo-electron microscopy data (Methods section ‘Model refinement’). Regions of the final electron density map are noticeably superior to the same regions generated using the Bragg peaks alone (Fig. 4). In particular, helices show better definition in side-chains, and the model better follows the electron density. The benefit of including the continuous diffraction data is quantified in the improvement of the  $R_{\text{free}}$  factor at low resolution (Extended Data Fig. 1). Although the assembled continuous

diffraction data are quite noisy, owing to the limited number (2,848) of diffraction patterns, the structure clearly benefits from their inclusion. Control experiments verified these benefits (see Methods section ‘Control analyses and comparisons’). Further improvements could be expected with more measurements.

Our approach, demonstrated with serial diffraction data collected using an XFEL, may also be applicable to larger crystals measured with conventional X-ray sources. The advantages of data collection with an XFEL include a higher tolerable dose and the capacity to measure data from hundreds of thousands of crystals, from which the most suitable can be selected solely on the basis of their diffraction patterns. The improved structural refinement at a resolution beyond that of measurable Bragg peaks establishes the origin of the continuous diffraction and demonstrates its utility in obtaining substantial increases in resolution without having to improve crystal lattice order. The crystal is simply a means to obtain an ensemble of oriented macromolecules to carry out aligned-molecule diffractive imaging<sup>30</sup>. We anticipate that the approach will be applicable to a range of macromolecular systems, and may allow *ab initio* phasing of crystals. After many decades of being discarded as of little use, all-too-common ‘poorly diffracting’ protein crystals might now be exploited for high-resolution structure determination.

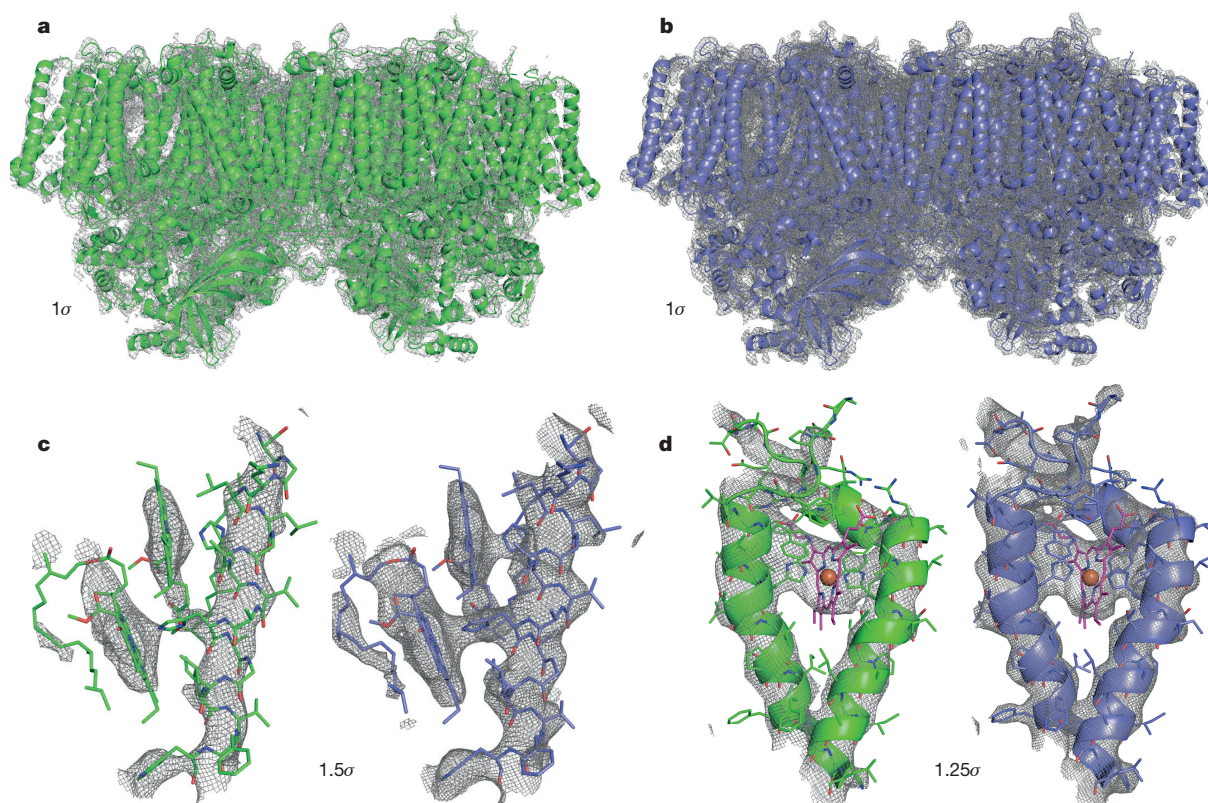




**Figure 3 | An unbiased size estimate of the rigid structural unit.**

This estimate is obtained by a Fourier transform of the continuous diffraction intensities, yielding the autocorrelation function (3D pair-distribution function). **a**, Projection of the experimentally determined 3D autocorrelation along the crystal *c* axis. **b**, **c**, The equivalent projections of the autocorrelation functions calculated from the 4.5-Å model of the PSII dimer (**b**) and the monomer (**c**) after applying the point-group symmetries

of the crystal. The extent of the rigid structural unit matches the size and shape of the PSII dimer. **d**, A loose support generated by thresholding and dilating the 4.5-Å-resolution structure was used as the support constraint for iterative phasing. **e**, Plots of the Fourier shell correlation (FSC; blue) and phase retrieval transfer function (PRTF; red). A value of PRTF = 0.5 is reached at a resolution of 3.5 Å. The dashed line corresponds to the boundary between the phased Bragg and the continuous diffraction.



**Figure 4 | Improvement in resolution and quality of electron density.**

This improvement is obtained by directly phasing the continuous transform using the method of coherent diffractive imaging. **a**, Electron density map of the PSII dimer after refinement using structure factors obtained only from Bragg peaks at 4.5-Å resolution, contoured at one standard deviation of the densities ( $1\sigma$ ). **b**, Electron density map obtained by iterative phase retrieval on the continuous diffraction data using the

support constraint (molecular envelope) of Fig. 3d at 3.5-Å resolution, contoured at  $1\sigma$ . **c**, **d**, Electron density maps of two regions of the PSII dimer, using only the Bragg diffraction (models shaded green), and the Bragg and continuous diffraction (models shaded blue). **c**, Two antenna chlorophylls in the antenna protein PsbC with their His residues, contoured at  $1.5\sigma$ . **d**, The haem group of PsbE/F, contoured at  $1.25\sigma$ .

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 13 July; accepted 14 December 2015.**

- Heras, B. & Martin, J. L. Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallogr.* **D61**, 1173–1180 (2005).
- Umena, Y., Kawakami, K., Shen, J.-R. & Kamiya, N. Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å. *Nature* **473**, 55–60 (2011).
- Sayre, D. Some implications of a theorem due to Shannon. *Acta Crystallogr.* **5**, 843 (1952).
- Wall, M. E., Adams, P. D., Fraser, J. S. & Sauter, N. K. Diffuse X-ray scattering to model protein motions. *Structure* **22**, 182–184 (2014).
- Fienu, J. R. Phase retrieval algorithms: a comparison. *Appl. Opt.* **21**, 2758–2769 (1982).
- Chapman, H. N. & Nugent, K. A. Coherent lensless X-ray imaging. *Nature Photon.* **4**, 833–839 (2010).
- Miao, J., Ishikawa, T., Robinson, I. K. & Murnane, M. M. Beyond crystallography: diffractive imaging using coherent x-ray light sources. *Science* **348**, 530–535 (2015).
- Guinier, A. *X-Ray Diffraction in Crystals, Imperfect Crystals, and Amorphous Bodies* Ch. 7 (Dover, 1994).
- Rupp, B. *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology* Ch. 7, 10 (Garland Science, 2009).
- Doucet, J. & Benoit, J. P. Molecular dynamics studied by analysis of the X-ray diffuse scattering from lysozyme crystals. *Nature* **325**, 643–646 (1987).
- Caspar, D. L. D., Clarage, J., Salunke, D. M. & Clarage, M. Liquid-like movements in crystalline insulin. *Nature* **332**, 659–662 (1988).
- Faure, P. et al. Correlated intramolecular motions and diffuse X-ray scattering in lysozyme. *Nature Struct. Biol.* **1**, 124–128 (1994).
- Pérez, J., Faure, P. & Benoit, J.-P. Molecular rigid-body displacements in a tetragonal lysozyme crystal confirmed by X-ray diffuse scattering. *Acta Crystallogr.* **D52**, 722–729 (1996).
- Wall, M. E., Ealick, S. E. & Gruner, S. M. Three-dimensional diffuse X-ray scattering from crystals of *Staphylococcal* nuclease. *Proc. Natl Acad. Sci. USA* **94**, 6180–6184 (1997).
- Welberry, T. R., Heerdegen, A. P., Goldstone, D. C. & Taylor, I. A. Diffuse scattering resulting from macromolecular frustration. *Acta Crystallogr.* **B67**, 516–524 (2011).
- Elser, V. & Millane, R. P. Reconstruction of an object from its symmetry-averaged diffraction pattern. *Acta Crystallogr.* **A64**, 273–279 (2008).
- Bates, R. H. T. Fourier phase problems are uniquely solvable in more than one dimension. I: underlying theory. *Optik* **61**, 247–262 (1982).
- Fienu, J. R. Reconstruction of a complex-valued object from the modulus of its Fourier transform using a support constraint. *J. Opt. Soc. Am. A* **4**, 118–123 (1987).
- Elser, V. Phase retrieval by iterated projections. *J. Opt. Soc. Am. A* **20**, 40–55 (2003).
- Miao, J., Charalambous, P., Kirz, J. & Sayre, D. Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. *Nature* **400**, 342–344 (1999).
- Robinson, I. K., Vartanyants, I. A., Williams, G. J., Pfeifer, M. A. & Pitney, J. A. Reconstruction of the shapes of gold nanocrystals using coherent X-ray diffraction. *Phys. Rev. Lett.* **87**, 195505 (2001).
- Chapman, H. N. et al. High-resolution *ab initio* three-dimensional X-ray diffraction microscopy. *J. Opt. Soc. Am. A* **23**, 1179–1200 (2006).
- Kupitz, C. et al. Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser. *Nature* **513**, 261–265 (2014).
- Chapman, H. N. et al. Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73–77 (2011).
- Boutet, S. et al. High-resolution protein structure determination by serial femtosecond crystallography. *Science* **337**, 362–364 (2012).
- White, T. A. et al. *CrystFEL*: a software suite for snapshot serial crystallography. *J. Appl. Cryst.* **45**, 335–341 (2012).
- Yefanov, O. et al. Mapping the continuous reciprocal space intensity distribution of X-ray serial crystallography. *Phil. Trans. R. Soc. B* **369**, 20130333 (2014).
- Shapiro, D. et al. Biological imaging by soft X-ray diffraction microscopy. *Proc. Natl Acad. Sci. USA* **102**, 15343–15346 (2005).
- Fischer, N. et al. Structure of the *E. coli* ribosome-EF-Tu complex at <3 Å resolution by Cs-corrected cryo-EM. *Nature* **520**, 567–570 (2015).
- Spence, J. C. H. & Doak, R. B. Single molecule diffraction. *Phys. Rev. Lett.* **92**, 198102 (2004).

**Acknowledgements** We acknowledge support of the Helmholtz Association through project-oriented funds and the Virtual Institute “Dynamic Pathways in Multidimensional Landscapes”; the DFG through the Gottfried Wilhelm Leibniz Program; the European Research Council under the European Union’s Seventh Framework Programme ERC Synergy Grant 609920 “AXSIS” and Marie Curie FP7-PEOPLE-2011-ITN Grant 317079 “Nanomem”; the BMBF through Project 05E13GU1; the Graduate College “GRK 1355” at the University of Hamburg, the International Max Planck Research School UFAST, the BioXFEL Science Technology Center (award 1231306); and the US National Institutes of Health (NIH), National Institute of General Medical Sciences grants R01 GM095583, U54 GM094599, and R01 GM097463. Parts of the sample injector used at the Linac Coherent Light Source (LCLS) for this research was funded by the NIH, P41GM103393, formerly P41RR001209. Use of the LCLS, SLAC National Accelerator Laboratory, is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under contract number DE-AC02-76SF00515.

**Author Contributions** H.N.C. conceived the idea of molecular imaging using imperfect crystals, and developed it with K.A., A.B., O.M.Y., D.O., V.M., K.R.B. and L.G. K.A., A.B., O.M.Y., D.O., L.G. and V.M. carried out the diffractive imaging analysis. P.F. led the photosystem II project. Serial femtosecond crystallography measurements were led by P.F. and A.B. and performed with S.R.-C., J.C., C.K., D.O., M.M., A.S., K.D., D.J., G.N., U.W., Y.Z. and M.S. M.L., A.A., M.S.H., J.S.R., J.E.K. and S.Bo. set up and operated the CXI instrument. S.Ba., R.F., A.B. and M.M. performed on-site data evaluation at LCLS. S.Ba. and R.F. provided the initial search model and carried out model refinement on the image with D.O. Photosystem II was isolated by S.R.-C., J.C., C.E.C., A.S., C.K., A.S., J.-H.Y. and K.D., from which microcrystals were grown by S.R.-C., J.C., C.E.C., A.S. and C.K. Test experiments on crystals were performed by K.D., I.S., P.L.X. and M.M. H.N.C., A.B., K.A., O.M.Y., D.O., L.G., P.F., J.C.H.S., T.A.W., S.Ba. and S.Bo. wrote the manuscript with improvements from all authors.

**Author Information** The coordinates of the atomic model obtained by molecular replacement from the Bragg peaks alone, and with the inclusion of continuous diffraction, have been deposited into the Protein Data Bank under accession codes 5E7C and 5E79, respectively. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.N.C. ([henry.chapman@desy.de](mailto:henry.chapman@desy.de)).



## METHODS

**Crystallography.** Microcrystals were grown by a free interface diffusion centrifugation technique, as described previously<sup>23</sup>. The size of the crystals ranged from 0.5  $\mu\text{m}$  to 5  $\mu\text{m}$  in width, which was determined by dynamic light scattering. The space group of the crystals is  $P2_12_12_1$  (ref. 23).

The experiments were carried out at the Coherent X-ray Imaging (CXI) end station<sup>31</sup> of the Linac Coherent Light Source (LCLS)<sup>32</sup>. X-ray pulses with durations of 40 fs and photon energy of 9.48 keV were focused using X-ray optics in a Kirkpatrick–Baez geometry to a beam size of about  $1 \times 1 \mu\text{m}^2$ . A suspension of PSII crystals in their crystallization buffer<sup>23</sup> was pre-filtered using 20- $\mu\text{m}$  stainless-steel filters and delivered at room temperature to the X-ray interaction region as a 3–4- $\mu\text{m}$ -diameter jet at a flow rate of  $10\text{--}15 \mu\text{L min}^{-1}$ , using a gas dynamic virtual nozzle (GDVN)<sup>33,34</sup> as previously described<sup>23</sup>. The average X-ray pulse energy at the sample was about 1.0 mJ, assuming a beamline transmission of 60%. The dose to each crystal was at most 275 MGy, and diffraction patterns were acquired on the basis of diffraction before destruction<sup>35,36</sup>. Diffraction frames were collected on a Cornell-SLAC Pixel Array Detector (CSPAD) at a rate of one frame per X-ray pulse, at 120 Hz (ref. 37). 1.24 million frames were collected, during an integrated collection time of 2.9 h.

The diffraction frames were initially processed following a standard data evaluation pipeline established for serial femtosecond crystallography. Frames containing crystal diffraction were found using Cheetah<sup>38</sup> on the basis of the identification of Bragg peaks. These ‘hits’, which accounted for about 5.0% of the frames, were transferred to CrystFEL (version 0.6.0)<sup>26,39</sup> along with the peak positions for automated indexing and to obtain a merged set of integrated Bragg intensities. Indexing the lattice determines the crystal orientation and crystal lattice parameters for each diffraction pattern, information that was later used to assemble the full 3D pattern shown in Fig. 2c. The relative coordinates of all pixels in the segmented CSPAD detector were determined by a procedure using a strongly diffracting standard (lysozyme)<sup>40</sup>. Using these calibrated pixel positions, a total of 25,585 patterns out of 61,946 were successfully indexed to create a final list of Bragg reflections and their integrated intensities. The resultant data set of Bragg intensities extended to a resolution of 4.5 Å, as indicated by the modified Pearson correlation coefficient CC\* (ref. 41) (Extended Data Fig. 2).

The crystal orientations obtained by indexing the lattice were subsequently used to assemble the individual diffraction-pattern snapshots into a 3D reciprocal-space grid representing the full-pattern intensity distribution of the average disordered crystal. The volume consisted of  $501 \times 501 \times 501$  cubic voxels, each with a width of  $1.364 \times 10^{-3} \text{ Å}^{-1}$  (where  $q = 1/d$ ). Thus, the magnitude of the wave-vector transfer  $q$  at the face centre of the voxel array was  $0.341 \text{ Å}^{-1}$ , corresponding to a resolution of 2.93 Å. Most of the individual snapshot patterns did not show noticeable continuous diffraction above a substantial featureless background caused by scattering from the water. The strength of the background depends linearly on the thickness of the liquid jet, and was removed by subtracting the average computed by summing counts along circles of constant  $q$  after correcting for the linear polarization of the X-ray beam. This radial average excluded very high and very low values to avoid the influence of Bragg peaks. The strength of the remaining crystal-dependent continuous diffraction signal for each frame was estimated by first replacing Bragg peak values with a local fit of the slowly varying continuous signal. This was achieved by using a thresholded median filter: pixel values that differed from a  $5 \times 5$ -pixel median-filtered copy of the pattern by more than one standard deviation above the mean were selected as bright pixels. The values of these, and pixels within a 3-pixel distance of them, were replaced with the values from the  $5 \times 5$ -pixel median-filtered copy. The strength of the continuous diffraction signal for each frame was best estimated from pixels in the range of  $0.1\text{--}0.2 \text{ Å}^{-1}$  after excluding the Bragg signal, and is plotted as a function of the total integrated Bragg-peak signal in Extended Data Fig. 3 for all of the 25,585 indexed patterns. The plot shows a strong linear correlation between these two signals. The incident intensity for each snapshot diffraction pattern fluctuates, owing to the random targeting of crystals to the X-ray focus profile and their variation in size. The trend shows that increased intensity or crystal size leads to an increase in both the Bragg and continuous diffraction. A single linear trend indicates little variation in the degree of lattice disorder of the crystals. We found that selecting the 2,848 patterns with the strongest continuous diffraction to create the full 3D pattern gave higher signal-to-noise and continuous-scattering contrast at high resolution than including all 25,585 patterns.

**The effect of rigid-body disorder on Bragg and continuous diffraction.** A deviation from strict translational correlation in a crystal gives rise to a component of the diffraction pattern that is continuous, which is usually studied in the context of protein dynamics<sup>4,10–14</sup>. The time-averaged diffraction of a moving object can be considered as the incoherent sum of static snapshots over the time of the exposure, and, for uncorrelated motions of structural units in a crystal, the time-averaged diffraction is equal to any particular snapshot in the limit of a large number of unit

cells. Thus, time-integrated diffraction measurements cannot distinguish static disorder from dynamic. In the case of XFEL exposures, the positions of structural units such as entire macromolecules are certainly static. Following the approach of ref. 42, we first derive the diffraction of a translationally disordered crystal containing only one rigid object per unit cell (that is,  $P1$  symmetry). The density of this crystal in terms of the real-space coordinate  $\mathbf{x}$ ,  $\rho_c(\mathbf{x})$ , can be described by

$$\rho_c(\mathbf{x}) = \rho_a(\mathbf{x}) \otimes \sum_{n=1}^N \delta(\mathbf{x} - \mathbf{a}_n - \Delta_n) \quad (1)$$

where  $\rho_a(\mathbf{x})$  is the density of the asymmetric unit, repeated over  $N$  positions in the crystal given by random displacement vectors  $\Delta_n$  from ideal lattice positions  $\mathbf{a}_n$ . In equation (1),  $\otimes$  represents a convolution and  $\delta(\mathbf{x})$  is the Dirac delta function. Without loss of generality, the mean displacement  $\langle \Delta_n \rangle = 0$ , and  $\langle \Delta_n^2 \rangle = \sigma^2$ .

The 3D diffraction intensity distribution is the squared modulus of the Fourier transform of the crystal electron density  $\rho_c(\mathbf{x})$ . This can also be obtained by calculating the Fourier transform of the electron-density autocorrelation function (the Patterson function),  $P(\mathbf{x}) = \rho_c(\mathbf{x}) \otimes \rho_c^*(-\mathbf{x})$ . For the disordered crystal

$$\begin{aligned} P(\mathbf{x}) &= \rho_a(\mathbf{x}) \otimes \sum_{n=1}^N \delta(\mathbf{x} - \mathbf{a}_n - \Delta_n) \otimes \rho_a^*(-\mathbf{x}) \otimes \sum_{k=1}^N \delta(\mathbf{x} + \mathbf{a}_k + \Delta_k) \\ &= \rho_a(\mathbf{x}) \otimes \rho_a^*(-\mathbf{x}) \otimes \left[ \sum_{n=1}^N \sum_{k=1}^N \delta(\mathbf{x} - \mathbf{a}_n - \Delta_n) \otimes \delta(\mathbf{x} + \mathbf{a}_k + \Delta_k) \right] \\ &= \rho_a(\mathbf{x}) \otimes \rho_a^*(-\mathbf{x}) \otimes \left[ \sum_{n=1}^N \sum_{k=1}^N \delta(\mathbf{x} - \mathbf{a}_n - \Delta_n + \mathbf{a}_k + \Delta_k) \right] \\ &= \rho_a(\mathbf{x}) \otimes \rho_a^*(-\mathbf{x}) \otimes \left[ N\delta(\mathbf{x}) + \sum_{n=1}^N \sum_{k=1, k \neq n}^N \delta(\mathbf{x} - \mathbf{a}_n - \Delta_n + \mathbf{a}_k + \Delta_k) \right] \end{aligned} \quad (2)$$

The double sum in equation (2) can be understood for large crystals in the following manner. A constant lattice difference vector  $\mathbf{a}_n - \mathbf{a}_k$  is obtained for many pairs of  $n$  and  $k$ , as illustrated in Extended Data Fig. 4. For this difference, the sum samples all possible displacement differences,  $\Delta_n - \Delta_k$ , except when  $n = k$ . Thus, the delta function at that lattice difference vector is broadened by the probability distribution of the displacement difference. If all unit cells are displaced independently, this probability distribution of the difference vectors will be the autocorrelation of the probability distribution of a single displacement. For a Gaussian with a standard deviation of  $\sigma$ , this is just another Gaussian with a standard deviation of  $\sqrt{2}\sigma$ . Thus

$$\begin{aligned} &\sum_{n=1}^N \sum_{k=1, k \neq n}^N \delta(\mathbf{x} - \mathbf{a}_n - \Delta_n + \mathbf{a}_k + \Delta_k) \\ &= \frac{\exp(-x^2/4\sigma^2)}{\sqrt{4\pi\sigma^2}} \otimes \sum_{n=1}^N \sum_{k=1, k \neq n}^N \delta(\mathbf{x} - \mathbf{a}_n + \mathbf{a}_k) \end{aligned} \quad (3)$$

Keeping in mind that the Fourier transform ( $\mathcal{F}$ ) maps convolutions to multiplications, the intensity can be calculated from equations (2) and (3) as

$$\begin{aligned} I(\mathbf{q}) &= |F_a(\mathbf{q})|^2 \mathcal{F} \left[ N\delta(\mathbf{x}) + \frac{\exp(-x^2/4\sigma^2)}{\sqrt{4\pi\sigma^2}} \otimes \sum_{n=1}^N \sum_{k=1, k \neq n}^N \delta(\mathbf{x} - \mathbf{a}_n + \mathbf{a}_k) \right] \\ &= |F_a(\mathbf{q})|^2 \mathcal{F} \left\{ N\delta(\mathbf{x}) \otimes \left[ \delta(\mathbf{x}) - \frac{\exp(-x^2/4\sigma^2)}{\sqrt{4\pi\sigma^2}} \right] \right. \\ &\quad \left. + \frac{\exp(-x^2/4\sigma^2)}{\sqrt{4\pi\sigma^2}} \otimes \sum_{n=1}^N \sum_{k=1}^N \delta(\mathbf{x} - \mathbf{a}_n + \mathbf{a}_k) \right\} \\ &= |F_a(\mathbf{q})|^2 \left[ N(1 - e^{-4\pi^2\sigma^2q^2}) + e^{-4\pi^2\sigma^2q^2} \sum_{n=1}^N \sum_{k=1}^N e^{2\pi i(\mathbf{a}_n - \mathbf{a}_k) \cdot \mathbf{q}} \right] \end{aligned} \quad (4)$$

where  $F_a(\mathbf{q})$  is the (complex-valued) Fourier transform of  $\rho_a(\mathbf{x})$  at a resolution of  $|q| = 1/d$ . Equation (4) shows that the diffraction pattern of a disordered crystal with a single asymmetric unit consists of the squared modulus of the molecular transform  $F_a(\mathbf{q})$ , modulated by two terms. The second term is the reciprocal lattice, which gives the Bragg peaks, further modulated by the Debye–Waller factor,  $\exp(-4\pi^2\sigma^2q^2)$ . The first term has no lattice associated with it, but is a monotonically increasing function with a value of 0 at  $q = 0$  that asymptotes to a constant



for  $q > 1/(2\pi\sigma)$ . Thus, at high  $q$  values, the diffraction is proportional to the continuous diffraction pattern of the rigid unit,  $|F_a(q)|^2$ .

Generalizing further to the case of  $M$  asymmetric units in a unit cell, one can write the unit cell electron density as

$$\sum_{m=1}^M \rho_a(R_m \mathbf{x} + \mathbf{t}_m)$$

where  $R_m$  and  $\mathbf{t}_m$  are the rotation and translation operators for the  $m$ th asymmetric unit. As in equation (1), the electron density for a disordered crystal is

$$\rho_c(\mathbf{x}) = \sum_{m=1}^M \sum_{n=1}^N \rho_a(R_m \mathbf{x} + \mathbf{t}_m) \otimes \delta(\mathbf{x} - \mathbf{a}_n - \Delta_{m,n})$$

where  $\Delta_{m,n}$  is the random displacement of the  $m$ th asymmetric unit in the  $n$ th unit cell.

The Patterson function of  $\rho_c(\mathbf{x})$  can be split into two terms, the first a sum of like terms connecting corresponding asymmetric units in different unit cells and the second consisting of the cross terms. Thus

$$\begin{aligned} P(\mathbf{x}) &= \sum_{m=1}^M \sum_{n=1}^N \rho_a(R_m \mathbf{x} + \mathbf{t}_m) \otimes \delta(\mathbf{x} - \mathbf{a}_n - \Delta_{m,n}) \otimes \sum_{j=1}^M \sum_{k=1}^N \rho_a^*(-R_j \mathbf{x} - \mathbf{t}_j) \\ &\quad \otimes \delta(\mathbf{x} + \mathbf{a}_k + \Delta_{j,k}) \\ &= \sum_{m=1}^M \left[ \rho_a(R_m \mathbf{x} + \mathbf{t}_m) \otimes \rho_a^*(-R_m \mathbf{x} - \mathbf{t}_m) \otimes \sum_n \delta(\mathbf{x} - \mathbf{a}_n - \Delta_{m,n}) \right. \\ &\quad \left. \otimes \sum_k \delta(\mathbf{x} + \mathbf{a}_k + \Delta_{m,k}) \right] \\ &\quad + \sum_{m=1}^M \sum_{j=1}^M \left[ \rho_a(R_m \mathbf{x} + \mathbf{t}_m) \otimes \rho_a^*(-R_j \mathbf{x} - \mathbf{t}_j) \otimes \sum_n \delta(\mathbf{x} - \mathbf{a}_n - \Delta_{m,n}) \right. \\ &\quad \left. \otimes \sum_k \delta(\mathbf{x} + \mathbf{a}_k + \Delta_{j,k}) \right] \end{aligned}$$

For the first term in this equation, the convolution of the two delta function sums can be evaluated in a similar way to that of the case of the single asymmetric unit in equation (3). This yields a term  $N\delta(\mathbf{x})$  for  $n = k$  and the blurred delta functions for all other cases. However, for the cross term, because different asymmetric units are assumed to move independently, this convolution does not have the  $N\delta(\mathbf{x})$  term. Taking the inverse Fourier transform of  $P(\mathbf{x})$  and simplifying yields

$$\begin{aligned} I(\mathbf{q}) &= \left[ \sum_{m=1}^M |F_a(R_m \mathbf{q})|^2 \right] \left[ N(1 - e^{-4\pi^2 \sigma^2 q^2}) + e^{-4\pi^2 \sigma^2 q^2} \sum_{n=1}^N \sum_{k=1}^N e^{2\pi i(\mathbf{a}_n - \mathbf{a}_k) \cdot \mathbf{q}} \right] \\ &\quad + \sum_{m=1}^M \sum_{j=1}^M F_a^*(R_m \mathbf{q}) F_a(R_j \mathbf{q}) e^{2\pi i \mathbf{q} \cdot (\mathbf{t}_m - \mathbf{t}_j)} \left[ e^{-4\pi^2 \sigma^2 q^2} \sum_{n=1}^N \sum_{k=1}^N e^{2\pi i(\mathbf{a}_n - \mathbf{a}_k) \cdot \mathbf{q}} \right] \end{aligned}$$

On gathering the Bragg terms and factorising, the intensity distribution simplifies to

$$\begin{aligned} I(\mathbf{q}) &= N \left[ \sum_{m=1}^M |F_a(R_m \mathbf{q})|^2 \right] (1 - e^{-4\pi^2 \sigma^2 q^2}) \\ &\quad + \left[ \sum_{m=1}^M F_a(R_m \mathbf{q}) e^{2\pi i \mathbf{q} \cdot \mathbf{t}_m} \right]^2 e^{-4\pi^2 \sigma^2 q^2} \sum_{n=1}^N \sum_{k=1}^N e^{2\pi i(\mathbf{a}_n - \mathbf{a}_k) \cdot \mathbf{q}} \end{aligned} \quad (5)$$

This equation has a similar form to equation (4), in that there are two terms, representing the continuous and Bragg diffraction. The difference is that each term is modulated by different combinations of the asymmetric unit transform. The continuous diffraction is the sum of the intensities of each asymmetric unit transform, whereas the Bragg peaks depend on the coherent sum of the asymmetric unit transforms (the unit-cell transform). This means that only the Bragg peaks are affected by the systematic absences of the space group, whereas the continuous diffraction depends only on the point group of the crystal.

This derivation started with the assumption of normally distributed displacements of rigid-body asymmetric units. The conclusions hold for other displacement distributions as well, whereby the term  $\exp(-4\pi^2 \sigma^2 q^2)$  is replaced by the Fourier transform of the autocorrelation of the probability distribution. For a smooth (band-limited) distribution, the Fourier transform will reduce to zero at high  $q$ , and the continuous diffraction term will dominate. Furthermore, the

derivation can be generalized to a number of non-identical rigid structural units in the unit cell.

**Two data sources from one experiment.** Structure determination could be considered in terms of a whole-pattern analysis, possibly based on equation (5). An alternative used here was to treat the Bragg diffraction and continuous diffraction as two separate data sources that both encode structural information of the macromolecular system. The observable Bragg diffraction covers the range  $0.03$ – $0.22 \text{ \AA}^{-1}$  ( $4.5$ – $\text{\AA}$  resolution). The continuous diffraction overlaps this range, but extends to a resolution of about  $2.2 \text{ \AA}$  (seen in the corners of Fig. 2d). However, to avoid the influence of the large pixel values of strong Bragg peaks when separating the continuous diffraction, we consider just the range  $0.22$ – $0.35 \text{ \AA}^{-1}$  for this data set. Our strategy for using both forms of diffraction data was to use the continuous diffraction to extend the resolution of a structure determined from the Bragg data alone and to add strong phasing constraints by directly accessing the molecular transform of the rigid structural unit.

Given equation (5), which predicts that the continuous diffraction is proportional to the incoherent sum of transforms of single rigid units aligned in their various crystallographic orientations, we expect the statistics of the continuous diffraction intensities to follow that of molecular diffraction, which should be exponentially distributed<sup>43</sup>. This can be seen in Extended Data Fig. 5, which shows a linear trend in the logarithmic histogram above a background level of around one count per voxel.

The size of the rigid structural unit that gives rise to the observed continuous diffraction pattern was estimated from the extent of the autocorrelation function, obtained by a Fourier transform of the continuous diffraction intensities. From equation (5), this yields the incoherent sum of autocorrelation functions of the structural unit in its various crystallographic orientations. To avoid the influence of the Bragg peaks on this calculation, this symmetrized autocorrelation function was generated from windowed regions of the full 3D pattern above a resolution of  $1/q = 4.5 \text{ \AA}$ , following a procedure previously described<sup>22</sup>. Using a Gaussian window function with a 15-pixel radius, 30 autocorrelation functions were generated in different locations in an orthogonal central section of the full 3D pattern. The functions produced from each subregion of reciprocal space were then averaged to obtain a single estimate of the projected symmetrized autocorrelation in the corresponding orthogonal direction (Fig. 3a). Although the windowing in reciprocal space reduces the spatial resolution of the determined autocorrelation map, it is still sufficient to identify the structural unit as an object of similar size to the PSII dimer. The assumption of the PSII dimer as the rigid structural unit was ultimately validated by the final electron density map obtained by phasing the continuous diffraction, and by improvements obtained in structural refinement (Extended Data Fig. 1), discussed below (another choice was also tested as described below). This result is not surprising because the dimer of PSII is the asymmetric unit of the crystals and is also the native biological oligomeric form, present in the photosynthetic membrane, maintained during the isolation procedure and in the process of crystallization.

The low-resolution structure from the Bragg data was obtained by molecular replacement phasing. The program F2MTZ was used to obtain an MTZ-file from the integrated intensities produced by CrystFEL. Structure-factor amplitudes were generated using TRUNCATE<sup>44</sup>. Molecular replacement phasing was performed using PHASER<sup>45</sup> in PHENIX<sup>46</sup> with the starting model being the PSII monomer from a published  $1.9$ – $\text{\AA}$ -resolution structure<sup>2</sup> (PDB, 3WU2). All ligands and co-factors except lipids, detergents, water and cryo-protectant molecules were kept in the input model. A composition containing the input model and its repeat was searched for, and PHASER found a single solution. Three rounds of simple refinement were performed using *phenix.refine* (with the program options *rigid body*, *XYZ coordinates* and *occupancies*). In the refinement, the 3WU2 structure was used as a reference model to restrain relationships between atoms to avoid overfitting. No further model building was carried out after this step. This process generated  $R$  and  $R_{\text{free}}$  values of  $24.8\%$  and  $27.2\%$ , respectively. A rendering of representative structural elements of PSII in the electron density map of the  $4.5$ – $\text{\AA}$ -resolution structure derived from Bragg peaks only is shown in Fig. 4 (the model rendered in green).

The  $4.5$ – $\text{\AA}$ -resolution electron density map was used to generate a binary map of the support of the PSII dimer (Fig. 3d) to be used for iterative phasing of the continuous diffraction. The inverse Fourier transform of the phased Bragg structure factors yields the unit-cell electron density map. To obtain the electron density of a single PSII dimer from the unit cell, *phenix.cut\_out\_density* was used to select only the density that was at most  $2 \text{ \AA}$  from any atom in the refined structural model of the PSII dimer. This density was Fourier transformed to obtain the squared modulus of the molecular transform up to  $1/q = 4.5 \text{ \AA}$ , which was then symmetrized by the point-group operations consisting of  $180^\circ$  rotations about each orthogonal axis and inversion through the origin (Extended Data Fig. 6). The similarity between this low-resolution computed continuous diffraction pattern

and the measurements gives further confidence in the choice of the PSII dimer as the rigid structural unit. To avoid over-constraining the phasing of the continuous diffraction data (described below), a dilated binary support mask was generated from the PSII dimer electron density by convolving it with a Gaussian of width 4.4 Å (three voxels) and thresholding the blurred density. Orthogonal slices through the support can be compared with the PSII dimer electron density image in Extended Data Fig. 7.

**Iterative phasing.** The 3D continuous diffraction was phased using the difference-map algorithm<sup>19</sup>. This algorithm iteratively refines the electron density image using an update rule that is expressed in terms of so-called projection operators operating in a finite-dimensional vector space. This  $N$ -dimensional space spans all possible complex-valued electron density images in a 3D array of  $N$  voxels in total. A vector  $\mathbf{x}$  in this space represents a particular electron density image. A projection of  $\mathbf{x}$  onto a constraint set  $C$ ,  $\mathcal{P}_C(\mathbf{x})$ , is the nearest point to  $\mathbf{x}$  that satisfies the constraint  $C$ . The difference map update rule is

$$\mathbf{x} \rightarrow \mathbf{x} + \mathcal{P}_D[2\mathcal{P}_S(\mathbf{x}) - \mathbf{x}] - \mathcal{P}_S(\mathbf{x})$$

where  $\mathcal{P}_S(\mathbf{x})$  is the support projection operator and  $\mathcal{P}_D(\mathbf{x})$  the data projection operator.

The support projection operator enforces the electron density to be zero outside the support volume (or low-resolution molecular envelope) of the PSII dimer described above. The data (or Fourier) projection operator enforces the electron density to be consistent with the measured Fourier magnitudes. In iterative phase retrieval of continuous coherent diffraction patterns, this projection operation is usually carried out by Fourier transforming the current iterate  $\mathbf{x}$  and setting the magnitude to the measured magnitude while keeping the phases from the iterate. Here, this projection was modified in two ways. First, the phases at low-resolution are provided by molecular replacement phasing, so, for these voxels in reciprocal space, the projection operator was set to simply replace the complex Fourier amplitude by these amplitudes and phases. Second, the point-group symmetries were applied to the squared Fourier magnitudes. This point-group symmetrization has been discussed in ref. 16. In the absence of symmetry, the multiplicative factor applied to the Fourier amplitudes is  $\sqrt{I_{\text{obs}} / |F_x|}$ , where  $F_x$  is the Fourier transform of the iterate and  $I_{\text{obs}}$  is the observed intensity. In the case of a symmetrized intensity, a similar factor is used, except that  $|F_x|$  is replaced by the point-group symmetrized version of itself.

Before applying the phasing algorithm, the molecular transform calculated from the phased Bragg intensities was scaled in magnitude and interpolated to match the merged continuous diffraction data. First, the low-resolution electron density grid was padded with zeros by an amount to approximately match the reciprocal-space voxel size to that of the continuous data. The zero-padded volume was then Fourier transformed and stretched using linear interpolation to make the voxel sizes match exactly. The relative scaling of the magnitude of continuous diffraction to Bragg diffraction depends on the lattice disorder length and detector pixel size. This scaling was determined empirically at an intermediate  $q$  range (corresponding to the resolution range 5.23–4.58 Å) in which both data types were non-zero and could be compared. The scale factor was chosen to minimize the sum of the squares of the intensity differences of the two data sets over the voxels in the resolution range.

Starting from a random white-noise guess for the voxel values of the electron density, the difference-map algorithm converged within 100 iterations, retrieving over 30 million phases. The convergence was achieved in far fewer iterations than is usual in 3D coherent diffractive imaging problems of similar size<sup>22</sup>, probably owing to the initial well-estimated support and foreknowledge of the low-resolution phases. The final 3D electron density image of the PSII dimer was calculated by averaging the complex amplitudes over another 100 iterations. Any variations in retrieved phases at a particular voxel will reduce the sum in that voxel as compared with the sum of the moduli. The ratio of these sums, known as the phase retrieval transfer function<sup>28</sup>, thus gives a measure of the consistency of the retrieved phases, and the complex sum naturally truncates the data where the phases are not known. In addition, this process was repeated with multiple random initial guesses. The solutions obtained were compared using a Fourier shell correlation (FSC) metric to confirm that there was no bias due to the choice of initial guess. Resolution-dependent plots of these metrics are shown in Fig. 3e, and the correlation between the continuous diffraction pattern and the symmetrized Fourier transform of the averaged iterate is shown in Extended Data Fig. 8a.

**Model refinement.** To interpret the 3.5-Å-resolution electron density volume image, and to compare with the lower-resolution refinement of the Bragg data, we refined a structural model to best match the diffractive image. Although it would make sense to refine a molecular model to the electron density map in real space, refinement was done in reciprocal space following a protocol recently established for interpreting cryo-electron microscopy maps using conventional crystallographic software<sup>29</sup>. The 3D electron density image (and hence the complex-valued

Fourier amplitudes) was taken to be a crystal of  $P1$  symmetry consisting of one PSII dimer in a lattice much larger than that of the real crystal. The phase probabilities, written in the form of Hendrickson–Lattman (HL) coefficients, were used for reciprocal-space refinement, which was performed against an MLHL target (maximum likelihood with experimental phase probability distribution) using PHENIX. Refinement statistics are shown in Extended Data Table 1.

**Control analyses and comparisons.** Two control analyses were performed to test whether the phasing of the continuous diffraction was an artefact of the input model. In the first test, the continuous diffraction was replaced by its orientational average, which is constant for a given  $q$  magnitude. In this case, any improvement in the quality of the reconstruction would be purely an artefact of the support constraint. We found that different random starts led to uncorrelated reconstructions, as shown by the FSC plot in Extended Data Fig. 8a. In the second test, the rigid structural unit was assumed to be the PSII monomer instead of the PSII dimer. Not only is this physically unexpected, owing to the stronger binding that forms the PSII dimer as compared with crystal contacts, but the symmetrized molecular transform of a monomer exhibits a larger speckle size and smaller extent of the autocorrelation than those observed (data not shown).

A comparison of the electron density maps of the molecular structure are given in Fig. 4 and Extended Data Fig. 9, in which improvements can be seen by including the continuous diffraction. In Extended Data Fig. 9, as well as showing the electron density obtained by iterative phasing of the continuous diffraction data, we show the calculated model density obtained by pseudo-crystal refinement. The correlation of the electron density maps<sup>47</sup> is given in Extended Data Fig. 9g, including a comparison with a previous high-resolution structure of ref. 2 truncated to a resolution of 3.5 Å.

**Constraint ratio.** The possibility to phase the diffraction pattern *ab initio*, such as by iterative methods, requires that the information content in the measurement exceeds that required to synthesize the electron density image. Such is not the case when measuring at Bragg peaks alone. The feasibility of phase retrieval in terms of a quantity  $\Omega$  which represents the ratio of independent measurements to the number of unknown coefficients required to describe the electron density of an object at a given resolution, is discussed in ref. 16. At a resolution of  $q = 1/d$ , the object can be represented in real space by samples spaced by distances of  $d/2$ . The number of unknown coefficients to describe the object is equal to the volume that the object occupies divided by the voxel volume  $d^3/8$ . The 3D region occupied by the object is known as the support of the object,  $S$ , with a volume given by  $V(S)$ . The number of independent Fourier-space measurements is likewise determined by the volume of the support of the autocorrelation function of the object, because the autocorrelation function is equivalent (through a Fourier transform) to the measured diffraction intensities. The autocorrelation function is Hermitian, because it is the Fourier transform of the real-valued intensities, and thus the number of independent coefficients is equal to half the volume of the support of the autocorrelation function  $V(A_S)$  divided by the same voxel volume  $d^3/8$ . Thus

$$\Omega = \frac{V(A_S)}{2V(S)}$$

In the absence of additional information, phase retrieval is feasible when  $\Omega > 1$ . As an example, consider an object that is shaped as a cuboid. The autocorrelation function is twice the length in each dimension, giving  $V(A_S) = 8V(S)$  and  $\Omega = 4$ .

The continuous diffraction produced by a disordered crystal is given by the incoherent sum of the transforms  $|F_a(R_m \mathbf{q})|^2$  for the orientations  $m$ , as described in equation (5). In this case, the number of independent measurements is equal to the volume of the union of the autocorrelation supports,  $V(\cup_m R_m A_S)$ , divided by the number of symmetry operations of the crystal point group. In the case of PSII, there are four molecular orientations. The support of the PSII dimer  $S$  is shown in Fig. 3d and one slice through the union of the autocorrelation supports  $\cup_m R_m A_S$  is given in Fig. 3b. The volume of this symmetrized support is 16.6 times larger than  $V(S)$ , yielding  $\Omega = 2.1$  when considering the four orientations and Hermitian symmetry. This value is lower than the case of only one orientation of the dimer, for which we find that the volume of the autocorrelation of the single support  $S$  of Fig. 3d is 10.9 times larger than  $V(S)$ , and so  $\Omega = 5.4$ . The effect of the symmetry is therefore not just a matter of counting symmetry operations, but depends on the shape of the autocorrelation function compared with the shape of the symmetrized autocorrelation. We see that even though the symmetry reduces the ratio  $\Omega$ , it is still greater than unity, indicating the feasibility of phase retrieval from the continuous diffraction data alone.

Compared with the continuous diffraction pattern of a single object, Bragg data contains fewer independent measurements, reducing the ratio  $\Omega$ . In a crystal, one certainly knows that the object is contained in a volume equal to that of the unit cell,  $V_{\text{cell}}$ . However, the autocorrelation function (known in this case as the Patterson function) is periodic with a unit-cell volume that is equal to that of the

crystal,  $V_{\text{cell}}$ . Thus, the Hermitian Patterson function contains half the number of independent coefficients as does the crystal unit cell itself, yielding  $\Omega = 1/2$ . This is Sayre's result<sup>2</sup>, and is eight times lower than for the single (non-crystalline) cuboid object mentioned above. Unlike the case of incoherent addition of the continuous diffraction of rigid units in the various crystallographic orientations, however, the constraint ratio for Bragg data does not depend on the space group. Although the unit-cell volume becomes larger to accommodate more molecules in a crystal of higher symmetry, this does not lead to a change in the information content even though it gives a higher density of Bragg peaks. For the space group  $P2_12_12_1$  of PSII, there are four asymmetric units per unit cell. Therefore, the number of unknowns is reduced by a factor of four. The number of independent measurements is reduced owing to the symmetry of the Patterson function. In this case, the symmetry of the Patterson function is  $Pmmm$ , which results in a factor of four reduction compared with  $P\bar{1}$ , and leads once more to  $\Omega = 1/2$ .

Non-crystallographic symmetry does help for Bragg data, as is well known for the method of molecular replacement, because it reduces the number of unknown coefficients needed to describe the object without necessarily reducing the number of independent measurements. A better knowledge of the support of the object also helps in this case, allowing the application of solvent flattening (equivalent to applying the support constraint in iterative phasing). We can calculate the constraint ratio  $\Omega$  for the PSII Bragg data, given the support  $S$  of Fig. 3d. In this case, the number of independent measurements is proportional to the volume of the unit cell, divided by four (the number of identical copies) and divided by two. Compared with the number of unknown coefficients, which is proportional to  $V(S)$ , we calculate

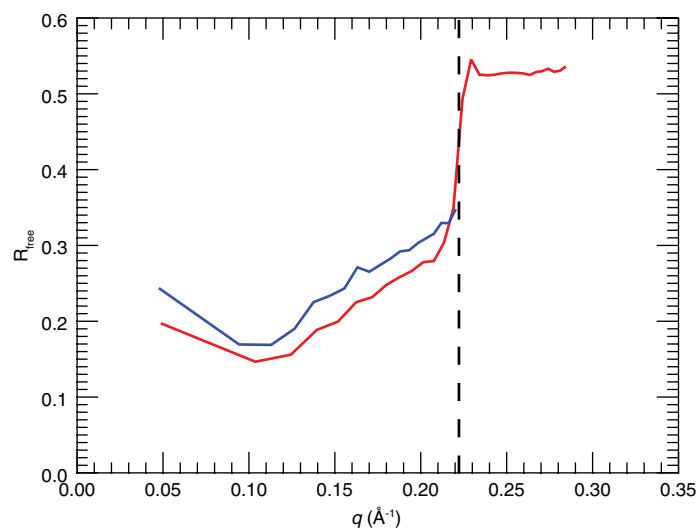
$$\Omega = \frac{V_{\text{cell}}}{8V(S)} = 0.86$$

where the unit-cell volume is determined from the lattice parameters provided in Extended Data Table 1. This value suggests that the phase retrieval problem for this crystal is not solvable with only the Bragg data.

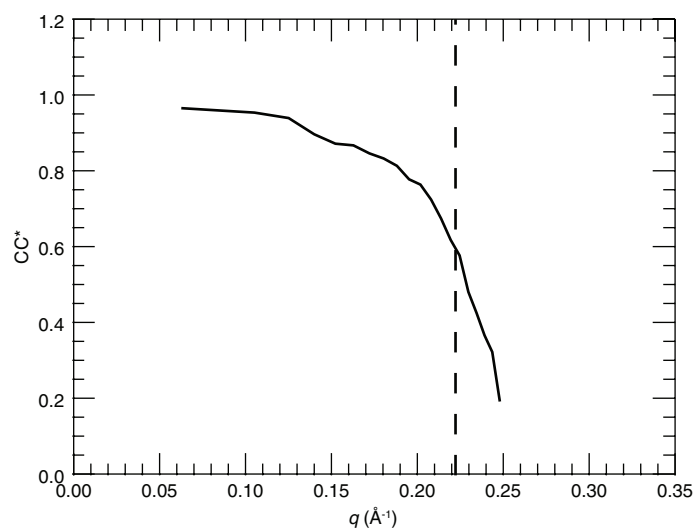
**Code availability.** The code used to perform the merging of diffraction patterns into a 3D volume and the iterative phasing of the continuous diffraction is available on request.

31. Liang, M. *et al.* The coherent X-ray imaging instrument at the Linac Coherent Light Source. *J. Synchrotron Rad.* **22**, 514–519 (2015).
32. Emma, P. *et al.* First lasing and operation of an ångström-wavelength free-electron laser. *Nature Photon.* **4**, 641–647 (2010).
33. DePonte, D. P. *et al.* Gas dynamic virtual nozzle for generation of microscopic droplet streams. *J. Phys. D* **41**, 195505 (2008).
34. Weierstall, U., Spence, J. C. H. & Doak, R. B. Injector for scattering measurements on fully solvated biospecies. *Rev. Sci. Instrum.* **83**, 035108 (2012).
35. Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752–757 (2000).
36. Barty, A. *et al.* Self-terminating diffraction gates femtosecond X-ray nanocrystallography measurements. *Nature Photon.* **6**, 35–40 (2012).
37. Hart, P. *et al.* The CSPAD megapixel x-ray camera at LCLS. In *X-Ray Free-Electron Lasers: Beam Diagnostics, Beamline Instrumentation, and Application* (eds Moeller, S. P. *et al.*) *Proc. SPIE* **8504**, 85040C (SPIE, 2012).
38. Barty, A. *et al.* *Cheetah*: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *J. Appl. Cryst.* **47**, 1118–1131 (2014).
39. White, T. A. *et al.* Crystallographic data processing for free-electron laser sources. *Acta Crystallogr.* **D69**, 1231–1240 (2013).
40. Yefanov, O. *et al.* Accurate determination of segmented X-ray detector geometry. *Opt. Express* **23**, 28459–28470 (2015).
41. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).
42. Cowley, J. M. *Diffraction Physics* 2nd edn, 150–151 (Elsevier, 1981).
43. Wilson, A. J. C. The probability distribution of X-ray intensities. *Acta Crystallogr.* **2**, 318–321 (1949).
44. French, G. S. & Wilson, K. S. On the treatment of negative intensity observations. *Acta Crystallogr.* **A34**, 517–525 (1978).
45. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
46. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr.* **D66**, 213–221 (2010).
47. Urzhumtsev, A., Afonine, P. V., Lunin, V. Y., Terwilliger, T. C. & Adams, P. D. Metrics for comparison of crystallographic maps. *Acta Crystallogr.* **D70**, 2593–2606 (2014).

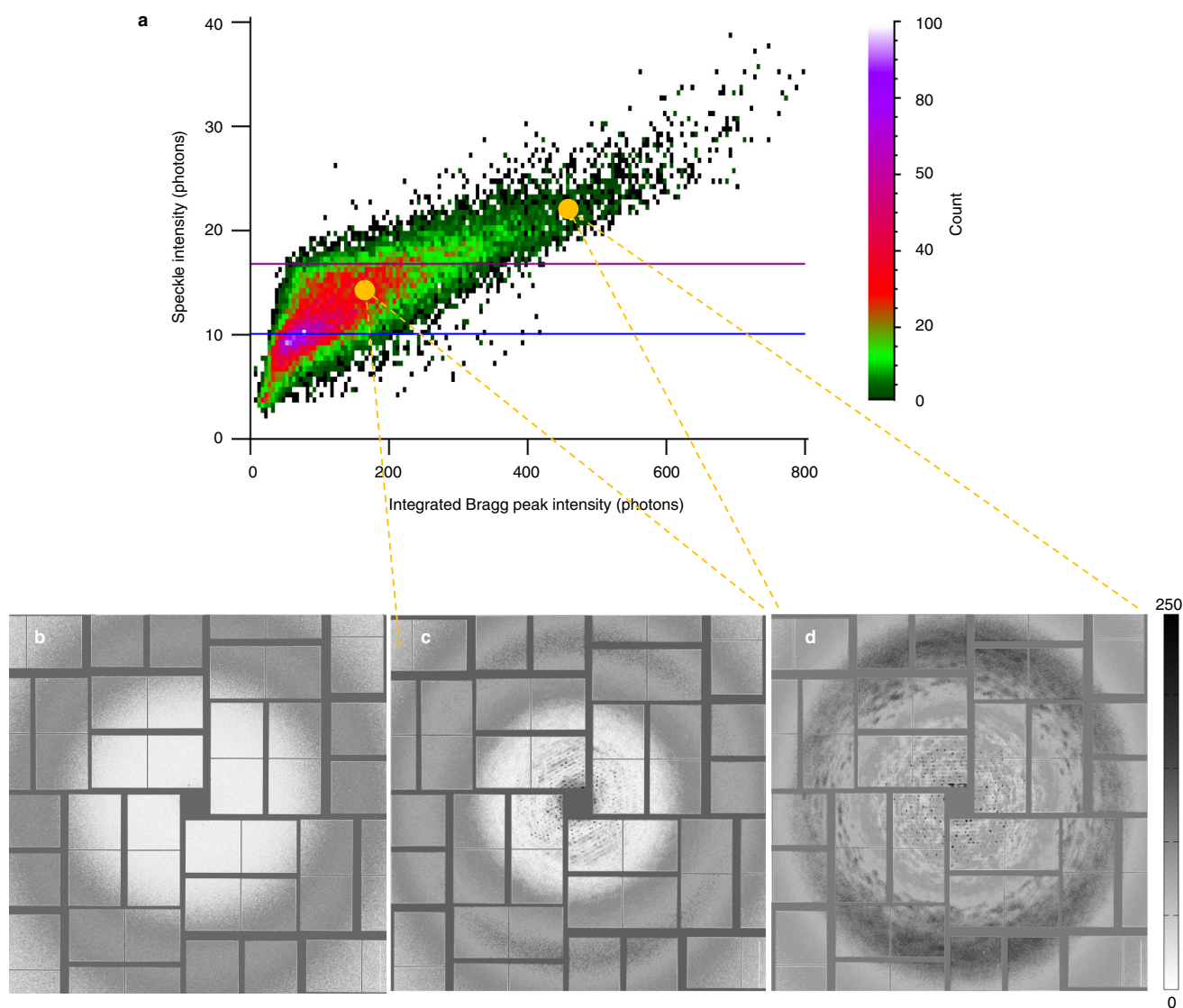




**Extended Data Figure 1 | Model refinement is improved at low resolution.** A plot of the metric  $R_{\text{free}}$  as a function of resolution shell  $q$ , showing a marked improvement of the model refined against the 3.5- $\text{\AA}$  diffractive image. The blue curve shows  $R_{\text{free}}$  prior to the inclusion of the continuous diffraction and the red curve shows  $R_{\text{free}}$  afterwards. Here  $R_{\text{free}}$  is calculated using only Bragg intensities (which were excluded from the refinement) for a resolution below 4.5  $\text{\AA}$ .



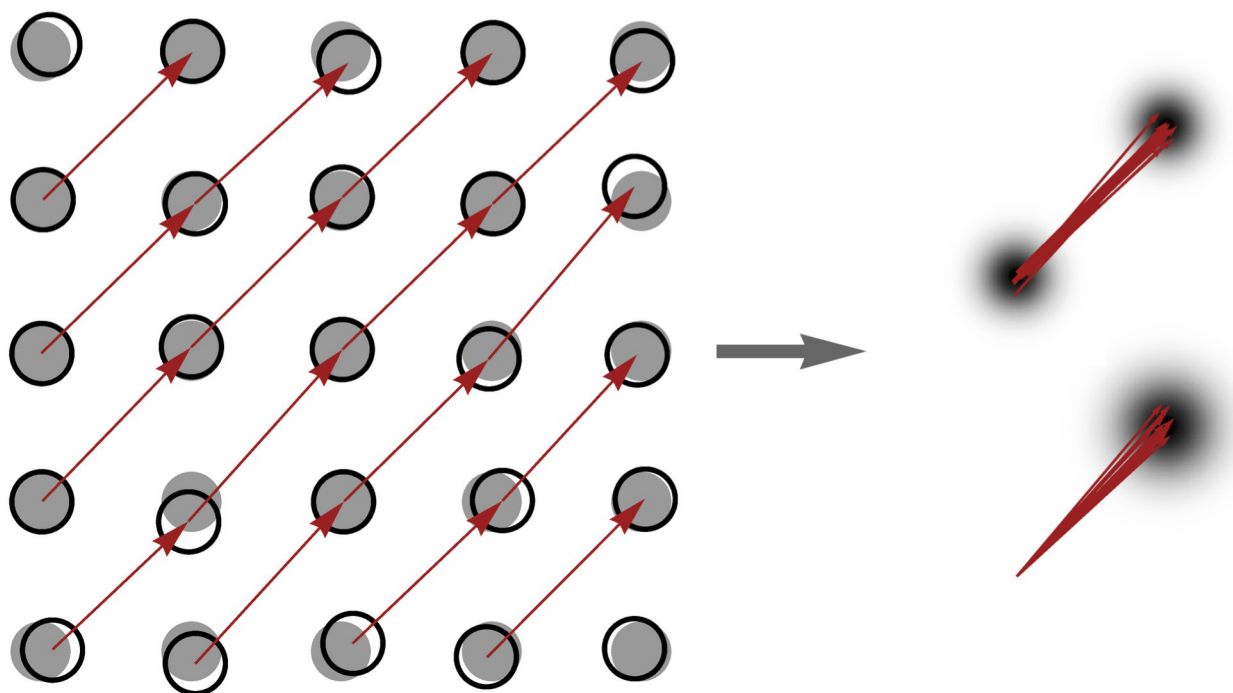
**Extended Data Figure 2 | Data quality and resolution of Bragg diffraction.** Plot of the reduced Pearson correlation coefficient,  $CC^*$  (ref. 41), as an estimate of the consistency of the integrated Bragg intensities determined from 25,585 indexed patterns. A value of  $CC^* = 0.5$  is reached at  $q = 0.23 \text{ \AA}^{-1}$ , or a resolution of  $4.3 \text{ \AA}$ .



**Extended Data Figure 3 | Strongest continuous diffraction occurs with strongest Bragg diffraction.** **a**, Two-dimensional histogram of patterns, sorted by the integrated counts in the continuous component of the diffraction pattern (in a  $q$  range of  $0.22\text{--}0.34\text{ \AA}^{-1}$ ) and the integrated signal in all detected Bragg peaks, for all 25,585 indexed patterns. We chose the 2,848 patterns with the strongest continuous diffraction signal above 17 X-ray counts (purple line) in the  $q$  region to generate the 3D

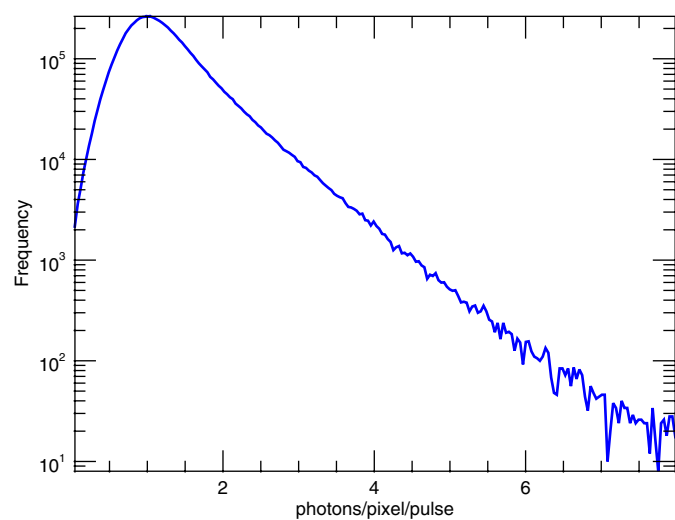
continuous pattern shown in Fig. 2c. The featureless background due to scattering from the solvent contributes 10 X-ray counts (blue line). **b**, The measurement from the liquid jet without crystals. **c**, **d**, Two representative patterns with speckle counts above the mean solvent background, but not above the threshold (**c**), and one of the 2,848 strongest patterns (**d**), as indicated by the positions of the yellow circles in **a**.



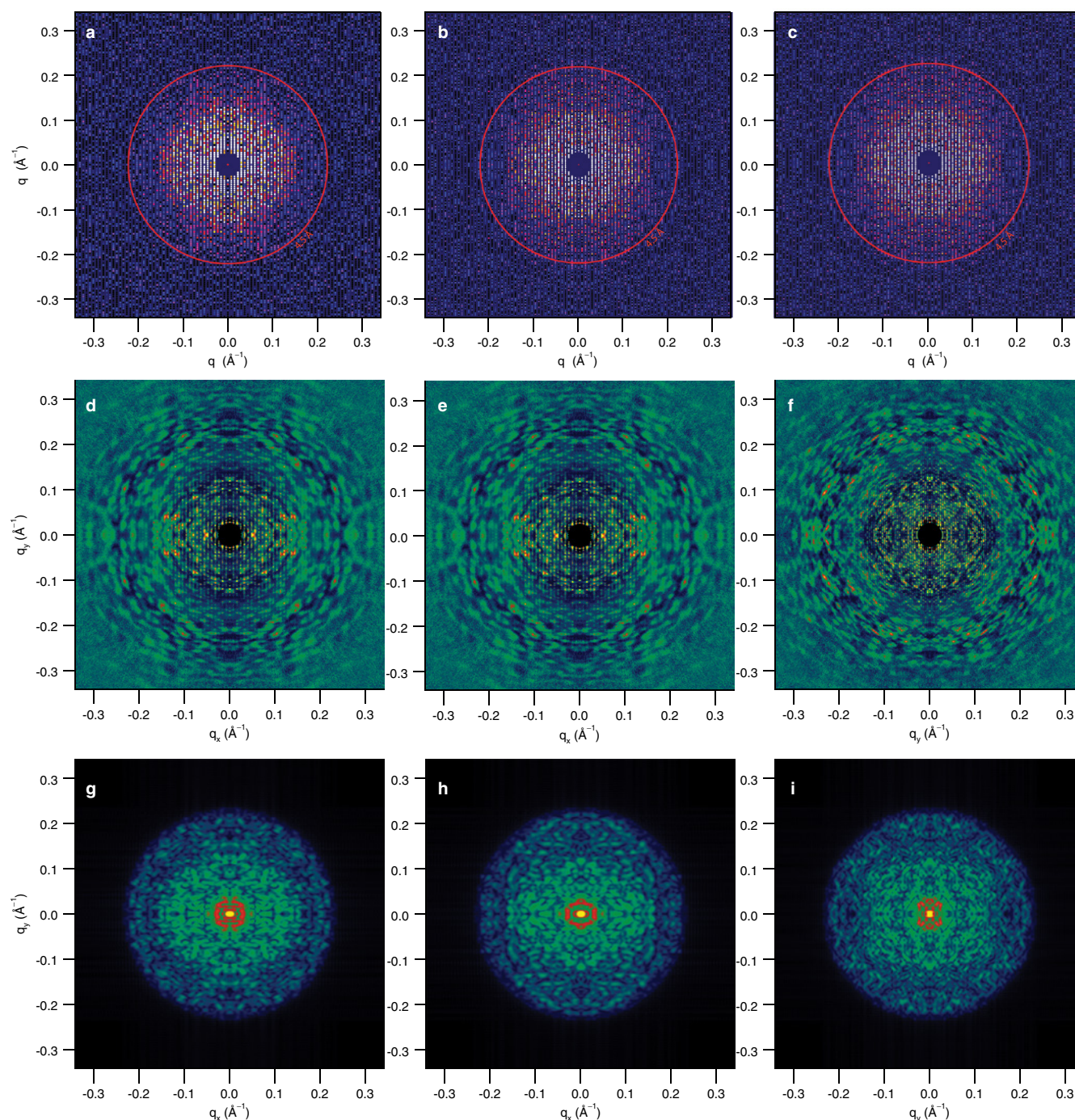


**Extended Data Figure 4 | Patterson function of a disordered lattice.** A distorted lattice (left, black circles, with ideal positions in grey), with vectors (red arrows) connecting all lattice points that have difference vectors  $\mathbf{a}_n - \mathbf{a}_k = (1, 1)$ . On the right, the arrows from the left panel are translated in two ways: upper right, the heads and tails are both displaced

from their ideal positions; bottom right, the tails are lined up, resulting in the distribution of head positions forming a broader Gaussian. In the limit of a large crystal, the resultant distribution is the autocorrelation of the displacement distribution. This process can be repeated for all difference vectors, leading to equation (3).



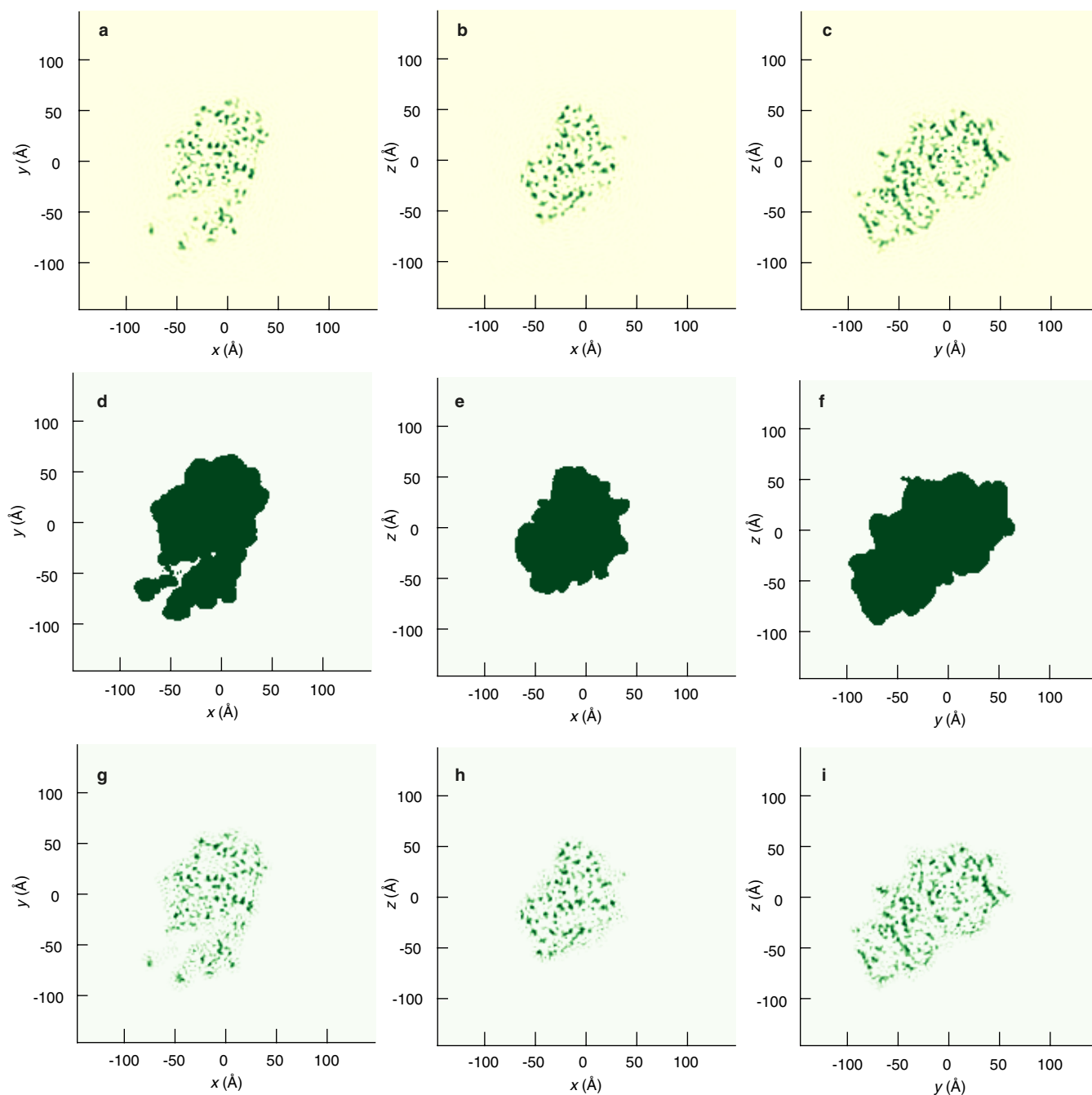
**Extended Data Figure 5 | Continuous diffraction exhibits Wilson statistics.** Histogram of merged continuous intensities in a  $q$  range of  $0.22\text{--}0.25\text{ \AA}^{-1}$ . Above a background level of around one photon per pixel per pulse, the logarithm of the histogram follows a linear trend with negative slope, characteristic of the exponential distribution predicted by Wilson statistics<sup>43</sup>.



**Extended Data Figure 6 | Central sections of the 3D full-pattern diffraction volume.** **a–c**, Bragg intensities in planes normal to the three orthogonal reciprocal-space axes,  $q_z$ ,  $q_y$  and  $q_x$ , respectively, which were arbitrarily chosen to be parallel to the  $c^*$ ,  $b^*$  and  $a^*$  axes of the PSII crystal. The 4.5-Å extent of the Bragg peaks is indicated by the red circle. **d–f**, Full-pattern diffraction intensities in central sections normal to the same three orthogonal axes as in **a–c**, respectively, obtained from that 2,848 strongest snapshot diffraction patterns. **g–i**, Continuous diffraction intensities calculated for a single PSII dimer using the model refined

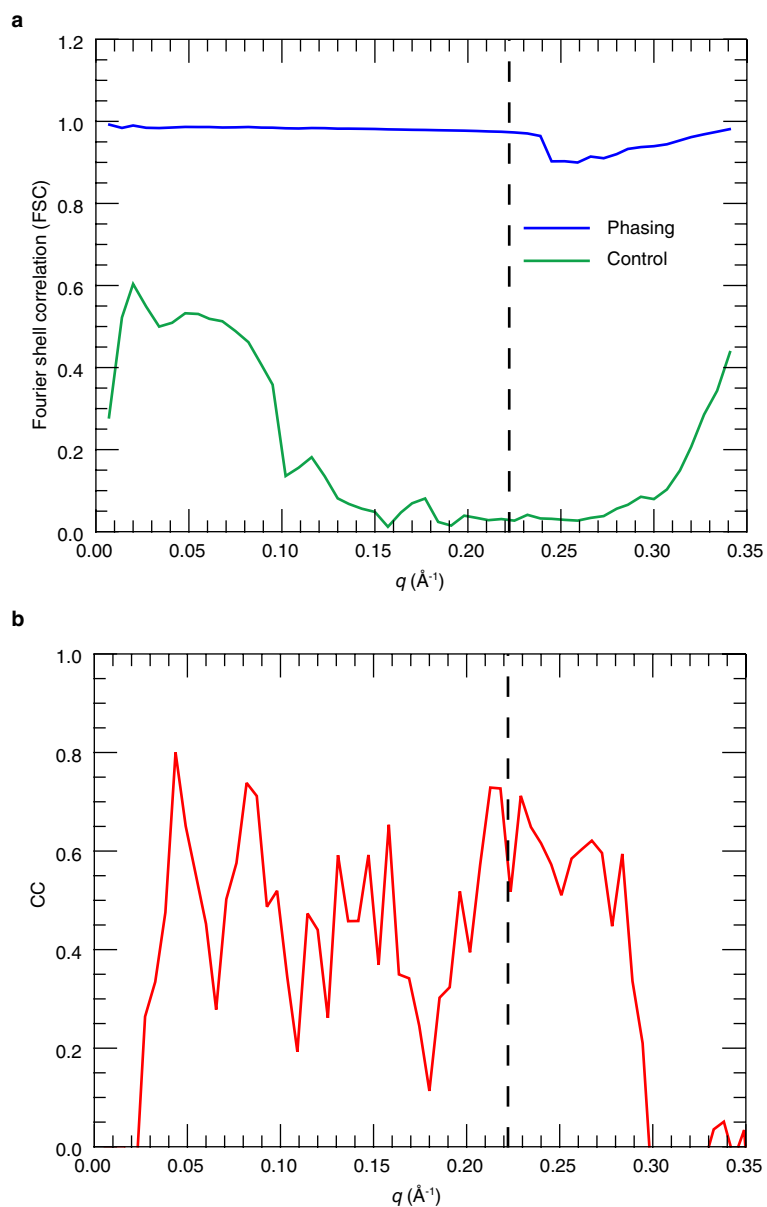
from the 4.5-Å Bragg data, for the same set of orthogonal planes as in **a–c**, respectively. The intensities were calculated from the incoherent sum of the squared modulus of the 3D molecular transform of a single (uncrystallized) PSII dimer in each of the four orientations of the 222 point group. All panels are plotted on the same scale, with the experimental data (**d–f**) extending to  $q = 0.33 \text{ \AA}^{-1}$  at the centre edge. The agreement between each of **d–f** with the corresponding panel **g–i** is further evidence that the rigid structural unit is the PSII dimer.





**Extended Data Figure 7 | Real-space orthogonal slices.** **a–c**, Slices of the observed structure factor ( $F_0$ ) electron density map, plotted as a grey scale, of a single PSII dimer obtained from the 4.5-Å Bragg intensities following model refinement of that data. Each slice is one pixel thick (1.5 Å) and is normal to the  $z$ ,  $y$  and  $x$  real-space axes, respectively (conjugate to  $q_z$ ,  $q_y$  and  $q_x$ ). **d–f**, Slices through the 3D real-space support constraint

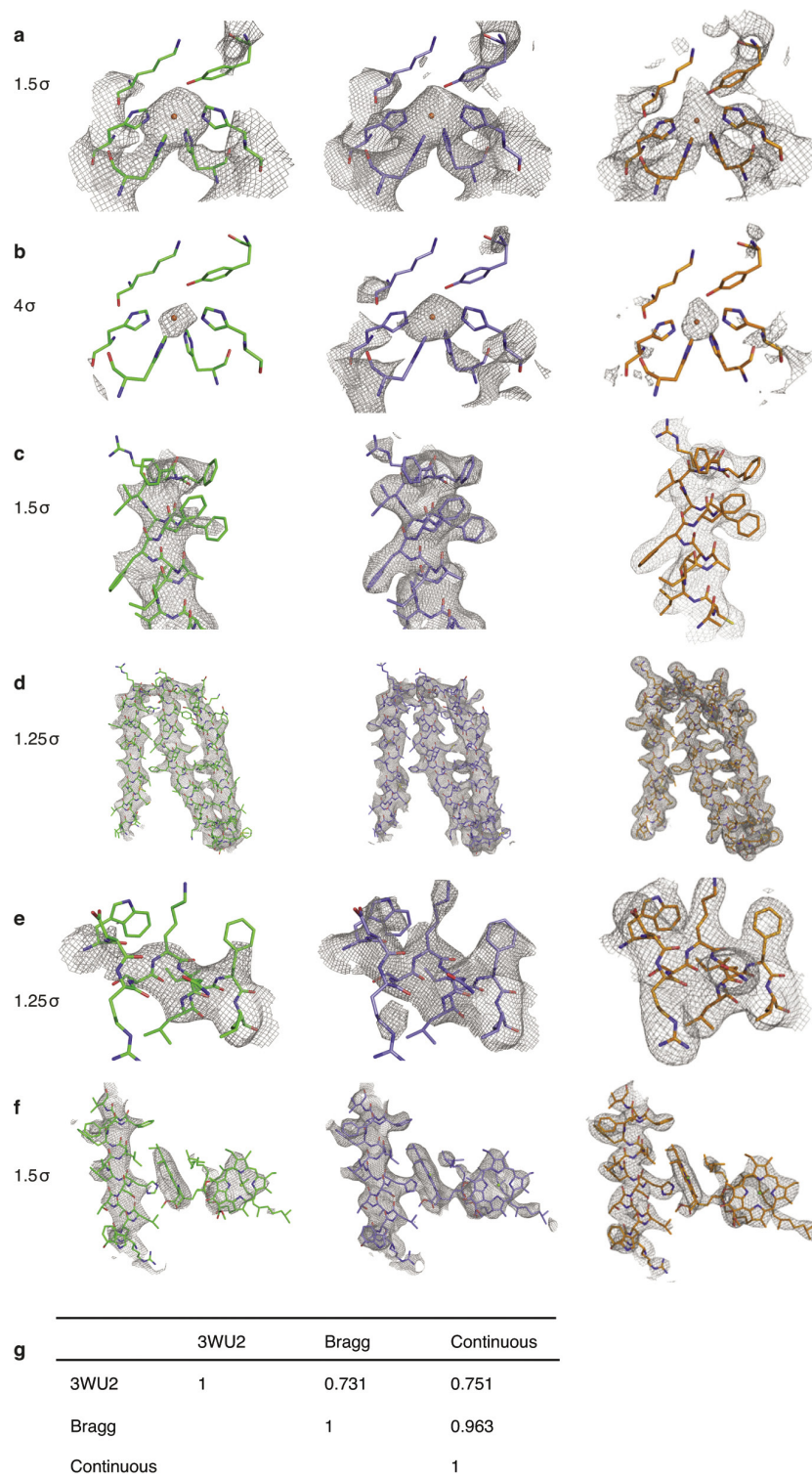
used for iterative phasing. The support was generated by blurring the 4.5-Å-resolution electron density map by 2.2 Å and then thresholding to achieve a binary mask. **g–i**, Slices through the 3.5-Å-resolution image obtained by iterative phasing of the continuous diffraction, using the support constraint illustrated in **d–f** and Fig. 3d.



**Extended Data Figure 8 | The continuous pattern is consistent with diffraction from a rigid object in crystallographic orientations.**

**a,** When the continuous diffraction intensities are substituted for intensities averaged over all orientations (that is, constant on surfaces of constant  $q$ ), iterative phasing using the support constraint of Fig. 3d fails, as indicated by this plot of the FSC as a function of resolution for solutions

(green) obtained from two independent phasing trials. The blue curve is the same FSC shown in Fig. 3e for comparison. The iterations never converged, so a phase retrieval transfer function for the control could not be generated. **b,** Plot of the cross correlation (CC) between the measured diffraction and that calculated from the determined electron density of a dimer, symmetrized by the four crystallographic orientations.



**Extended Data Figure 9 | Electron density maps of regions of the PSII dimer.** **a–f**, Electron density maps based on maximum likelihood structure factors ( $2mF_o - DF_c$ ) obtained using the Bragg diffraction (model shaded green, left), the Bragg and continuous diffraction (model shaded blue, middle) and computed from pseudo-crystal refinement (model shaded brown, right). The maps are rendered at various density levels relative to the standard deviation of the overall density. **a, b**, Non-haem iron coordinated by two His residues from D1 (chain A) and two from D2 (chain D), contoured at 1.5σ (**a**) and 4σ (**b**). Neighbouring Tyr and Lys residues are displayed as well. **c**, Part of an  $\alpha$ -helix (chain T), showing that the side chains (for example, Arg and Phe) fit better in the electron density when applying our new method (maps contoured at 1.5σ). **d**, Helices from

chains Y and Z (maps contoured at 1.25σ). The density map shows more details and better agreement to the model when applying the analysis using the Bragg and continuous diffraction. **e**, Detailed view of a section of chain Z (maps contoured at 1.25σ). Using only the Bragg diffraction, no electron density is visible at this level around the side chains of Trp, Lys and Arg. Again, the model fits better into the map when using the continuous diffraction. **f**, Two chlorophylls and part of the transmembrane helix of chain C (maps contoured at 1.5σ). **g**, Matrix of Pearson correlation coefficients of electron density maps obtained from the model presented in ref. 2 (PDB, 3WU2), the model refined from Bragg diffraction and that obtained from the continuous diffraction.



Extended Data Table 1 | Data collection, phasing and refinement

	Bragg diffraction	Continuous diffraction	Pseudo-crystal refinement
Point Group	222	222	1
Space Group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>		P1
<i>a</i> , <i>b</i> , <i>c</i> (Å)	133.25, 226.26, 307.09	735.3, 735.3, 735.3	250.8 250.8 250.8
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Resolution range (Å)	30–4.5 (4.62–4.5)	30–3.20 <sup>‡</sup>	30–3.5
Wilson <i>B</i> (Å <sup>2</sup> )	191.6		
Indexed / Oriented patterns	25,585	2,848	
Completeness (%)	99.9		
<i>I</i> /sig( <i>I</i> )	5.01 (1.49)		
R-split (%)	11.48 (46.16)		
CC*	0.99 (0.94)		
Maximum dose (MGy)	275		
<b>Refinement and phasing</b>			
Method	ML*	Difference Map	MLHL <sup>†</sup>
No. reflections / phases	55,609	$3.5 \times 10^7$	769,253
Resolution range (Å)	30.0–4.50 (4.58–4.50)	30.0–3.20 <sup>‡</sup>	A: 30.0–3.50 (3.54–3.50) B: 30.0–4.50 (4.55 – 4.50)
Rwork/Rfree (%)	24.8 / 27.2 (34.9 / 36.2)		A: 31.7 / 32.4 (53.2 / 54.2) B: 21.7 / 23.0 (39.6 / 40.6)
No. atoms	50,074		50,074
No. voxels		$1.26 \times 10^8$	
Bond lengths (Å)	0.005		0.012
Bond angles (°)	1.400		1.817

\*ML, maximum likelihood.

<sup>†</sup>MLHL, maximum likelihood combined with experimental phase probability distribution.<sup>‡</sup>3.20-Å resolution in the centre edge of the 3D voxel array.

# Covariation of deep Southern Ocean oxygenation and atmospheric CO<sub>2</sub> through the last ice age

Samuel L. Jaccard<sup>1,2</sup>, Eric D. Galbraith<sup>3,4,5</sup>, Alfredo Martínez-García<sup>6,7</sup> & Robert F. Anderson<sup>8</sup>

**No single mechanism can account for the full amplitude of past atmospheric carbon dioxide (CO<sub>2</sub>) concentration variability over glacial–interglacial cycles<sup>1</sup>. A build-up of carbon in the deep ocean has been shown to have occurred during the Last Glacial Maximum<sup>2,3</sup>. However, the mechanisms responsible for the release of the deeply sequestered carbon to the atmosphere at deglaciation, and the relative importance of deep ocean sequestration in regulating millennial-timescale variations in atmospheric CO<sub>2</sub> concentration before the Last Glacial Maximum, have remained unclear. Here we present sedimentary redox-sensitive trace-metal records from the Antarctic Zone of the Southern Ocean that provide a reconstruction of transient changes in deep ocean oxygenation and, by inference, respired carbon storage throughout the last glacial cycle. Our data suggest that respired carbon was removed from the abyssal Southern Ocean during the Northern Hemisphere cold phases of the deglaciation, when atmospheric CO<sub>2</sub> concentration increased rapidly, reflecting—at least in part—a combination of dwindling iron fertilization by dust and enhanced deep ocean ventilation. Furthermore, our records show that the observed covariation between atmospheric CO<sub>2</sub> concentration and abyssal Southern Ocean oxygenation was maintained throughout most of the past 80,000 years. This suggests that on millennial timescales deep ocean circulation and iron fertilization in the Southern Ocean played a consistent role in modifying atmospheric CO<sub>2</sub> concentration.**

Ice-core and surface-ocean proxy records provide a wealth of evidence supporting a central role of the Southern Ocean in modulating the air–sea partitioning of carbon during ice ages<sup>4,5</sup>. The Southern Ocean could exert a substantial control on the partial pressure of CO<sub>2</sub> ( $p_{\text{CO}_2}$ ), owing to its leverage on the efficiency of the global soft-tissue pump<sup>2</sup> (STP) by which the photosynthetic production, sinking and remineralization of organic matter store dissolved inorganic carbon (DIC) in the ocean interior. At present, vertical exchange of water causes deeply sequestered CO<sub>2</sub> and nutrients to be brought rapidly to the Southern Ocean surface, while iron (Fe) limitation of phytoplankton prevents the exposed nutrients from being entirely fixed back to organic matter before the water sinks again, allowing a net release of CO<sub>2</sub> to the atmosphere. Either decreasing the rate of vertical exchange or enhancing export production would raise DIC concentrations in deep waters, lowering atmospheric CO<sub>2</sub>. Higher DIC, in turn, enhances deep water corrosivity, dissolving carbonate (CaCO<sub>3</sub>) minerals and thereby increasing ocean alkalinity, which produces a further CO<sub>2</sub> drawdown<sup>2</sup>.

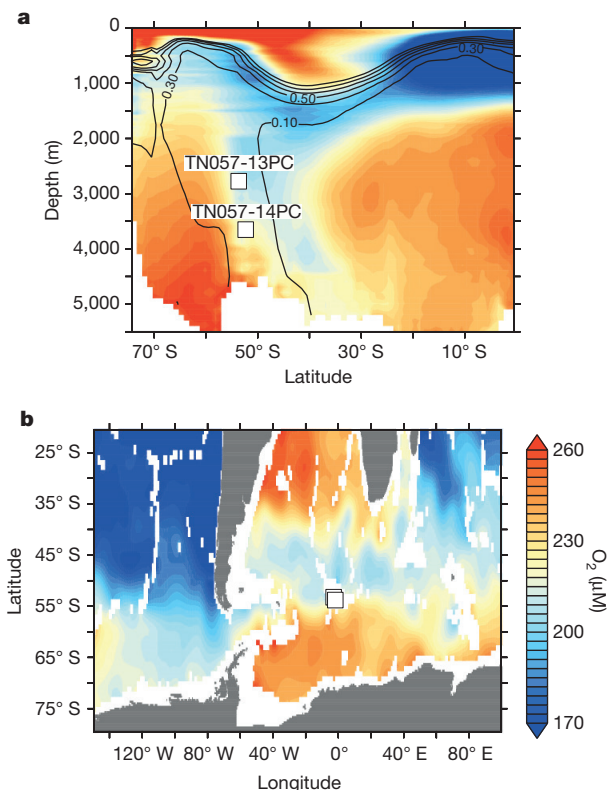
Phytoplankton productivity in the Subantarctic Zone of the Southern Ocean surface was stimulated during glacial periods by an enhanced supply of Fe-bearing dust, which would have reduced the leakage of CO<sub>2</sub> to the atmosphere, strengthening the overall STP<sup>6,7</sup>. Meanwhile, radiocarbon measurements on deep-sea corals and benthic foraminifera have been interpreted as showing weakened exchange between the

surface and the deep Southern Ocean during the Last Glacial Maximum (LGM)<sup>3,8</sup>; if this is correct, such a weakened exchange would have complemented the effect of dust<sup>9</sup>. From the LGM to the Holocene, the combination of dwindling dust inputs and accelerated vertical exchange in the Southern Ocean would have released carbon from the deep sea<sup>5</sup>, contributing to the rise of atmospheric CO<sub>2</sub>. However, as yet, very little evidence has been reported for STP-driven CO<sub>2</sub> storage in the deep Southern Ocean that extends before the LGM.

Here we report new geochemical observations of changes in the oxygenation of the deep Southern Ocean, chronicling STP carbon storage over the last 80,000 years (80 kyr). DIC storage by the STP, as defined here, is equal to zero in the ocean's surface layer, even though surface waters can contain large amounts of residual carbon inherited from deep waters owing to the slow air–sea exchange of CO<sub>2</sub>. This non-equilibrium component in surface waters is referred to as 'disequilibrium DIC' (ref. 10) and would be affected by sea ice, whereas STP DIC storage is independent of sea ice cover. Vertical stable carbon isotope ( $\delta^{13}\text{C}$ ) gradients in the Southern Ocean show large changes on glacial–interglacial and millennial (that is, 1–5 kyr) timescales<sup>11</sup>. Although these gradients are influenced by storage of respiratory CO<sub>2</sub> in the deep ocean, they are also susceptible to large disequilibrium effects that may have decoupled  $\delta^{13}\text{C}$  from DIC storage, since the air–sea equilibration of carbon isotopes is still an order of magnitude slower than for DIC itself<sup>12</sup>. Because dissolved oxygen (O<sub>2</sub>) is consumed stoichiometrically during the respiration of sinking organic matter, while being replenished during air–sea exchange an order of magnitude faster than CO<sub>2</sub>, it can provide a robust constraint on STP-induced carbon storage.

New measurements were made on two sediment cores raised from 2,800 m and 3,600 m depth at sites just to the north of the Weddell Sea (Methods, Fig. 1), where the bulk of Antarctic Bottom Water (AABW) forms through shelf processes and deep convection<sup>13</sup>. Modern observations show the presence of man-made chlorofluorocarbon-11 (CFC-11) near the core site, unambiguously identifying waters recently ventilated at the Southern Ocean surface (Fig. 1a). Idealized simulations with a fully coupled ocean–atmosphere–biogeochemistry model show that when deep convection in the Weddell Sea is weakened, waters at the core locations become depleted in oxygen, even in a simulation for which North Atlantic Deep Water is somewhat enhanced (Extended Data Fig. 1). At the same time, any change in export production and remineralization rates, including Fe fertilization by abundant glacial dust mobilized from South America during ice ages, would have altered the sinking flux of organic matter to the deep Southern Ocean, thereby changing the consumption rate of oxygen. Prior reconstructions have shown that glacial export fluxes were higher only in the Subantarctic Zone of the Southern Ocean<sup>14</sup>, so that the Antarctic Zone sites explored here would have been affected by the southward mixing of oxygen-depleted waters from lower latitudes. Thus, the core sites lie at a critical nexus of the Southern Ocean, where the impact of both deep southern

<sup>1</sup>Institute of Geological Sciences, University of Bern, Bern, Switzerland. <sup>2</sup>Oeschger Center for Climate Change Research, University of Bern, Bern, Switzerland. <sup>3</sup>Department of Earth and Planetary Sciences, McGill University, Montreal, Canada. <sup>4</sup>Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain. <sup>5</sup>Institut de Ciència i Tecnologia Ambientals and Department of Mathematics, Universitat Autònoma de Barcelona, Barcelona, Spain. <sup>6</sup>Geological Institute, ETH Zurich, Zurich, Switzerland. <sup>7</sup>Climate Geochemistry Department, Max Planck Institute for Chemistry, Mainz, Germany. <sup>8</sup>Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York, USA.

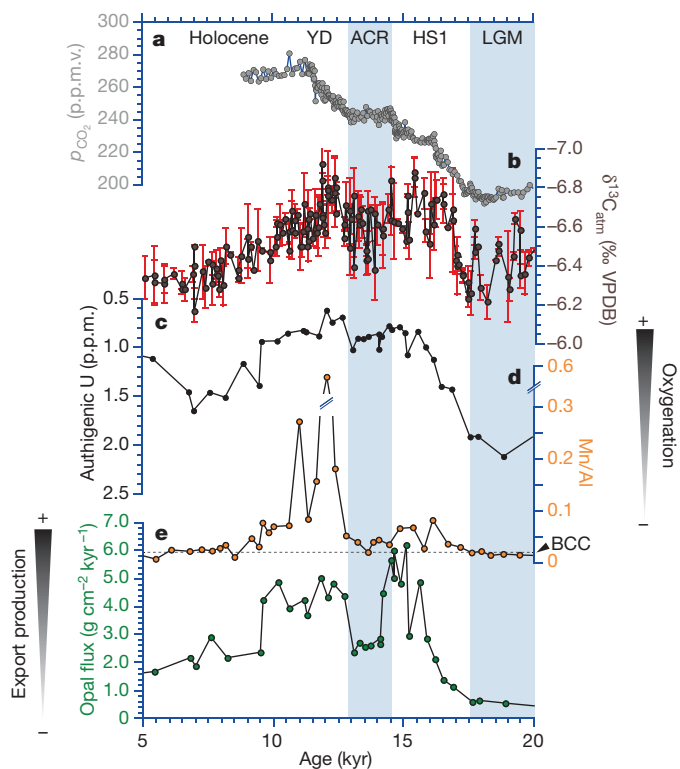


**Figure 1 | Modern oceanographic context.** **a, b**, Present-day dissolved oxygen concentration<sup>26</sup> is shown on a meridional transect of the South Atlantic, averaged between 25° W and 10° E (**a**) and on a horizontal surface at a water depth of 2,800 m (**b**). Contours in **a** show the concentration of chlorofluorocarbon-11 (CFC-11), in picomoles per kilogram (ref. 27). Both core site locations are bathed by well oxygenated, recently ventilated subsurface water masses and observations suggest that the two sedimentary archives are sensitive recorders with which to monitor changes in deep ocean oxygenation.

ventilation and Fe fertilization on the STP can be recorded by the sedimentary redox state.

The fundamental control on sedimentary redox conditions is the balance between the rate of supply of oxygen from overlying bottom waters and the rate of pore-water oxygen removal associated with respiration, which is regulated by the supply of organic matter. Comparing the sedimentary distribution of redox-sensitive trace metals with biogenic fluxes to the sea floor thus permits these processes to be disentangled, so that relative changes in oxygen concentrations at the water–sediment interface can be inferred. Manganese (Mn) enrichments are typically preserved in well oxygenated sediments, whereas dissolved uranium (U) typically diffuses across the water–sediment interface and subsequently precipitates from pore waters under more reducing conditions (Methods). These mineral precipitates are referred to as authigenic phases. Monitoring the sedimentary distribution of these two metals thus provides sensitivity across the full range of oxygen concentrations typically encountered in the open ocean.

At site TN057-13PC, authigenic U concentrations are highest during peak glacial conditions (LGM in Fig. 2c) and decrease rapidly thereafter, coinciding with the increase in opal flux (Fig. 2e), a proxy for the rain of organic material to the seabed<sup>12</sup> (Extended Data Fig. 2). Higher authigenic U during intervals of lower organic carbon supply to the sea floor can only reasonably be the result of decreased bottom-water oxygen concentrations<sup>15,16</sup>, as also observed in prior low-resolution measurements (Extended Data Fig. 3 and Extended Data Table 1), indicating increased storage of remineralized carbon in the deep Southern Ocean during the LGM. Given that the solubility of oxygen in seawater is greater in cold waters, the observation that the deep



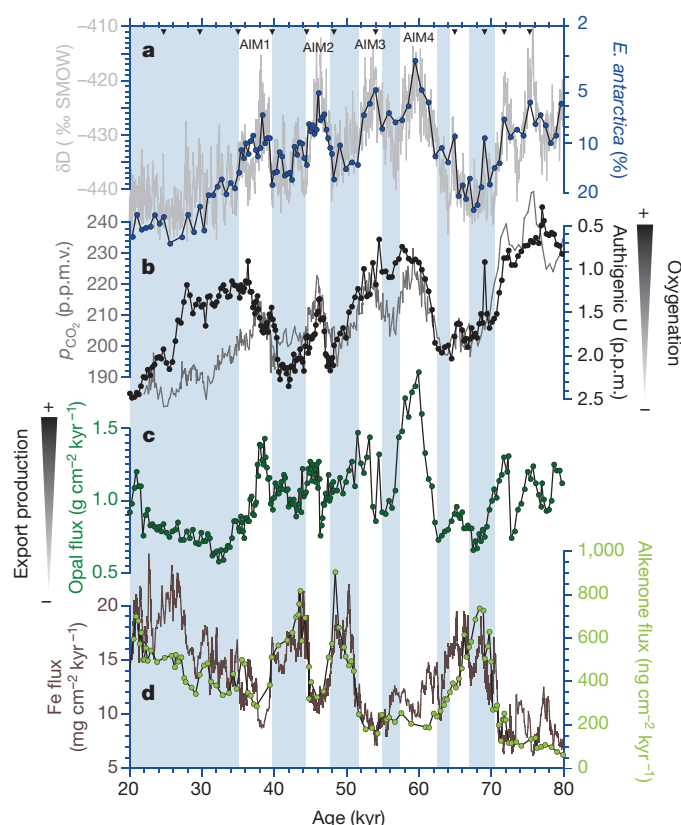
**Figure 2 | Palaeoclimatic reconstructions covering the last glacial termination at site TN057-13PC compared with ice core records.**

**a, b**,  $p_{\text{CO}_2}$  (**a**; ref. 28) and atmospheric  $\delta^{13}\text{C}$  (**b**; ref. 4, error bars represent the standard deviation of replicate measurements) from Antarctic ice cores. **c**, Sedimentary authigenic U concentrations (reversed scale). **d**, Sedimentary Mn/Al. **e**,  $^{230}\text{Th}$ -normalized biogenic opal flux<sup>29</sup>. Shaded intervals highlight cold periods in the Southern Hemisphere. YD, Younger Dryas; ACR, Antarctic Cold Reversal; HS1, Heinrich Stadial 1. The stippled line represents the bulk continental crust (BCC) Mn/Al background value (0.0167; ref. 30). The error ( $1\sigma$ ) related to the determination of authigenic U is smaller than the symbol size. p.p.m., parts per million; p.p.m.v., parts per million by volume; VPDB, Vienna Pee-Dee belemnite standard.

Atlantic was less well oxygenated during the LGM indicates that the solubility effect on oxygenation was overwhelmed by larger changes in the accumulated utilization of oxygen in the ocean interior<sup>10</sup>.

Subsequently, during Heinrich Stadial 1 (HS1, 17.5–14.7 kyr ago), authigenic U accumulation decreased (Fig. 2c) while Mn enrichments rose above the detrital background (Fig. 2d), indicating improved oxygenation of bottom water. This is in striking contrast to the change observed in the deep North Atlantic, where oxygen concentrations fell to their lowest values during HS1, reflecting the reduced input of oxygenated North Atlantic Deep Water<sup>17</sup>. Increased oxygenation during HS1 must have included the effect of decreasing organic matter flux in the Subantarctic Zone, given the rapidly dwindling Fe supply to the surface ocean<sup>6</sup>, but the opposing behaviour with the North Atlantic is also consistent with a north–south ventilation seesaw<sup>18</sup>. Importantly, a strong Mn enrichment recurred during the Younger Dryas (12.8–11.5 kyr), when the changes in Fe supply were too small to have altered deep water oxygenation much<sup>7</sup> (Extended Data Fig. 4). Thus, greater oxygenation of deep water at this location during the Younger Dryas event provides evidence for an increased proportion of oxygen-rich AABW relative to oxygen-depleted circumpolar deep water. A northward shift of the circumpolar fronts may have contributed to this increased oxygenation, but if spatial gradients of dissolved oxygen were similar to those that exist today (Fig. 1) then more rapid formation of AABW must have occurred as well, to induce the inferred increase in oxygen. Together, the records imply two phases of STP weakening in the Southern Ocean: the first (HS1) included combined effects of dust





**Figure 3 | Palaeoclimatic reconstructions covering the interval MIS2 to MIS5a at site TN057-14PC and Ocean Drilling Program (ODP) site 1090 (42.9° S, 8.9° E, 3,700 m) compared with ice core records.**

**a**, Relative abundance of diatom *Eucampia antarctica* (blue)—a sea surface temperature proxy—in TN057-14PC (ref. 29) compared to the Antarctic ice core local temperature reconstruction (light grey)<sup>31</sup> on the new AICC2012 age scale<sup>32</sup>. **b**, Ice-core atmospheric  $p_{\text{CO}_2}$  reconstructions<sup>33,34</sup> (dark grey) and sedimentary authigenic U concentrations at site TN057-14PC (black, reversed scale). **c**,  $^{230}\text{Th}$ -normalized biogenic opal flux at site TN057-14PC (ref. 29). **d**,  $^{230}\text{Th}$ -normalized Fe (brown) and alkenone (green) fluxes at ODP site 1090 (ref. 6). Triangles indicate the tiepoints used to generate the new version of the TN057-14PC age scale. Shaded intervals highlight cold intervals in the Southern Hemisphere, whereas unshaded intervals correspond to North Atlantic cold periods (Antarctic Isotope Maximum (AIM) 1–4 in the Southern Hemisphere). The error ( $1\sigma$ ) related to the determination of authigenic U is smaller than the symbol size. SMOW, Vienna mean ocean water standard.

supply and a north–south seesaw of ventilation, and probably reflected most of the net glacial–interglacial change<sup>10</sup>, while the second (Younger Dryas) was dominated by a transient pulse of ventilation alone. These deglacial observations are consistent with the stable carbon isotopic composition of atmospheric  $\text{CO}_2$  ( $\delta^{13}\text{C}_{\text{atm}}$ ) (ref. 4) (Fig. 2b) in suggesting that carbon was removed from the deep Southern Ocean during two episodes of improved deep ocean oxygenation that occurred during HS1 and the Younger Dryas.

Prior to the LGM, atmospheric  $\text{CO}_2$  concentrations rose repeatedly by 10–20 parts per million during periods of intense cold conditions in the North Atlantic, while Antarctica warmed. These intervals correspond to perturbations of the Atlantic Meridional Overturning and advances of sea ice in the North Atlantic, similar to HS1 and the Younger Dryas<sup>19</sup>, when deep North Atlantic  $\text{O}_2$  concentrations plummeted<sup>17</sup>. Our new data show that each of these intervals was also marked by decreasing authigenic U concentrations (Fig. 3b), concurrent with a greater rain of organic detritus to the sea floor (Fig. 3c). Thus, similar to the millennial-timescale changes of the last glacial termination, the oxygenation proxies are consistent with the release of deeply sequestered remineralized carbon to the atmosphere during

climate oscillations between 35 kyr and 80 kyr ago. Despite uncertainties in age models and the potential for nonlinear behaviour in authigenic U accumulation, the Pearson's correlation coefficient  $r^2$  between the TN057-14PC authigenic U and the ice-core  $\text{CO}_2$  is 0.63 ( $P = 10^{-19}$ ,  $n = 89$ ) over the period 35–80 kyr ago. We note less consistency between 25 kyr and 35 kyr ago, when authigenic U was lower than expected from the  $p_{\text{CO}_2}$  level; we were unable to find any sedimentological artefact that might have introduced a bias over this interval (Extended Data Fig. 5), and therefore suggest that at this time, during which there was a major expansion of the Northern Hemisphere ice sheets<sup>20</sup>,  $p_{\text{CO}_2}$  was lowered by a mechanism unrelated to the Southern Ocean's control on the STP, although this result should be further tested.

Our results show that carbon storage in the deep Southern Ocean was generally amplified when Antarctica was cold, consistent with arguments based on the similarity of Antarctic temperature and atmospheric  $\text{CO}_2$  over glacial cycles. However, the results also highlight a fundamental difference between the glacial–interglacial transition, during which the  $\text{O}_2$  changes in the north and south deep Atlantic correlate, and the millennial-timescale changes, for which  $\text{O}_2$  changes are antiphased between north and south. This difference can be explained by a contrast in the dynamics of ocean circulation, which were dominated by the bipolar ventilation seesaw on millennial timescales<sup>21</sup> rather than the steady-state adjustment to different  $\text{CO}_2$  levels, orbital forcing and ice sheet configuration that determined the glacial–interglacial change<sup>10</sup>. Iron fertilization, caused by enhanced dust input during Antarctic cold phases<sup>6</sup>, would have contributed some degree of the observed Southern Ocean STP strengthening during both LGM and millennial-timescale changes. In this light, it may seem remarkable that changes in dust fluxes and deep ventilation were coordinated (Extended Data Table 2), reinforcing their synergistic impacts on the STP.

We suggest that the coordination between ventilation and Fe supply on millennial timescales occurred through a mechanistic link between the two. Model simulations have suggested that a weakening of the Atlantic Meridional Overturning Circulation can lead to enhanced Southern Ocean ventilation via direct ocean circulation changes<sup>21</sup> as well as by forcing a migration in the Southern westerly winds<sup>22</sup>. In addition, it has been shown that a southward shift in the precipitation-bearing Southern Westerlies could initiate the retreat of Patagonian glaciers<sup>23</sup>. We suggest that, owing to this coupling of ocean and atmospheric circulation, millennial-timescale periods of enhanced deep Southern Ocean ventilation generally co-occurred with southward shifts of the Southern Westerlies, which simultaneously caused Patagonian glaciers to retreat, reducing both the production of fine-grained lithogenic material by glacial erosion<sup>24</sup> and the supply of dust by trapping glacial flour in proglacial lakes<sup>25</sup>. Thus, shifts in atmospheric circulation, coupled with changes in the Atlantic Meridional Overturning Circulation, would have caused the intensity of Fe limitation and upwelling to co-vary on millennial timescales, while the intensity of deep convection was altered by both atmospheric and oceanic drivers. These coupled processes, coordinating the input flux of carbon to the deep sea as well as its release by deep convection and upwelling, can explain the remarkable power of the Southern Ocean in modifying the efficiency of carbon storage over millennial timescales, and thereby atmospheric  $\text{CO}_2$  levels.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 12 March; accepted 27 November 2015.**

**Published online 3 February 2016.**

1. Brovkin, V., Ganopolski, A., Archer, D. & Munhoven, G. Glacial  $\text{CO}_2$  cycle as a succession of key physical and biogeochemical processes. *Clim. Past* **8**, 251–264 (2012).
2. Sigman, D. M., Hain, M. P. & Haug, G. H. The polar ocean and glacial cycles in atmospheric  $\text{CO}_2$  concentration. *Nature* **466**, 47–55 (2010).
3. Skinner, L. C., Fallon, S., Waelbroeck, C., Michel, E. & Barker, S. Ventilation of the deep Southern Ocean and deglacial  $\text{CO}_2$  rise. *Science* **328**, 1147–1151 (2010).

4. Schmitt, J. *et al.* Carbon isotope constraints on the deglacial CO<sub>2</sub> rise from ice cores. *Science* **336**, 711–714 (2012).
5. Martínez-Botí, M. A. *et al.* Boron isotope evidence for oceanic carbon dioxide leakage during the last deglaciation. *Nature* **518**, 219–222 (2015).
6. Martínez-García, A. *et al.* Iron fertilization of the Subantarctic Ocean during the last ice age. *Science* **343**, 1347–1350 (2014).
7. Anderson, R. F. *et al.* Biological response to millennial variability of dust and nutrient supply in the Subantarctic South Atlantic Ocean. *Phil. Trans. R. Soc. Lond. A* **372**, 20130054 (2014).
8. Burke, A. & Robinson, L. F. The Southern Ocean's role in carbon exchange during the last deglaciation. *Science* **335**, 557–561 (2012).
9. Jaccard, S. L. *et al.* Two modes of change in Southern Ocean productivity over the past million years. *Science* **339**, 1419–1423 (2013).
10. Galbraith, E. D. & Jaccard, S. L. Deglacial weakening of the oceanic soft tissue pump: global constraints from sedimentary nitrogen and oxygenation proxies. *Quat. Sci. Rev.* **109**, 38–48 (2015).
11. Ziegler, M., Diz, P., Hall, I. R. & Zahn, R. Millennial-scale changes in atmospheric CO<sub>2</sub> levels linked to the Southern Ocean carbon isotope gradient and dust flux. *Nature Geosci.* **6**, 457–461 (2013).
12. Galbraith, E. D., Kwon, E. Y., Bianchi, D., Hain, M. P. & Sarmiento, J. L. The impact of atmospheric pCO<sub>2</sub> on carbon isotope ratio of the atmosphere and ocean. *Glob. Biogeochem. Cycles* **29**, 307–324 (2015).
13. de Lavergne, C., Palter, J. B., Galbraith, E. D., Bernardello, R. & Marinov, I. Cessation of deep convection in the open Southern Ocean under anthropogenic climate change. *Nature Clim. Change* **4**, 278–282 (2014).
14. Kohfeld, K., Die Quéré, C., Harrison, S. P. & Anderson, R. F. Role of marine biology in glacial-interglacial CO<sub>2</sub> cycles. *Science* **308**, 74–78 (2005).
15. François, R. *et al.* Contribution of Southern Ocean surface-water stratification to low atmospheric CO<sub>2</sub> concentrations during the last glacial period. *Nature* **389**, 929–935 (1997).
16. Frank, M. *et al.* Similar glacial and interglacial export bioproductivity in the Atlantic sector of the Southern Ocean: multiproxy evidence and implications for glacial atmospheric CO<sub>2</sub>. *Paleoceanography* **15**, 642–658 (2000).
17. Hoogakker, B. A. A., Elderfield, H., Schmiedl, G., McCave, I. N. & Rickaby, R. E. M. Glacial–interglacial changes in bottom-water oxygen content on the Portuguese margin. *Nature Geosci.* **8**, 40–43 (2014).
18. Skinner, L. C., Waelbroeck, C., Scrivner, A. E. & Fallon, S. J. Radiocarbon evidence for alternating northern and southern sources of ventilation of the deep Atlantic carbon pool during the last deglaciation. *Proc. Natl Acad. Sci. USA* **111**, 5480–5484 (2014).
19. Böhm, E. *et al.* Strong and deep Atlantic meridional overturning circulation during the last glacial cycle. *Nature* **517**, 73–76 (2014).
20. Lambeck, K., Roubey, H., Purcell, A., Sun, Y. & Sambridge, M. Sea level and global ice volumes from the Last Glacial Maximum to the Holocene. *Proc. Natl Acad. Sci. USA* **111**, 15296–15303 (2014).
21. Schmittner, A., Brook, E. J. & Ahn, J. in *Ocean Circulation: Mechanisms and Impacts – Past and Future Changes of Meridional Overturning* Vol. 173 *Geophysical Monograph Series* (eds Schmittner, A., Chiang, J. & Hemming, S. R.) 392 (American Geophysical Union, 2007).
22. Chiang, J. C. H., Lee, S.-Y., Putnam, A. E. & Wang, X. South Pacific Split Jet, ITCZ shifts, and atmospheric north-south linkages during abrupt climate changes of the last glacial period. *Earth Planet. Sci. Lett.* **406**, 233–246 (2014).
23. Boex, J. *et al.* Rapid thinning of the late Pleistocene Patagonian Ice Sheet followed migration of the Southern Westerlies. *Sci. Rep.* **3**, 2118 (2013).
24. Herman, F. *et al.* Worldwide acceleration of mountain erosion under a cooling climate. *Nature* **504**, 423–426 (2013).
25. Sugden, D. E., McCulloch, R. D., Bory, A. J.-M. & Hein, A. S. Influence of Patagonian glaciers on Antarctic dust deposition during the last glacial period. *Nature Geosci.* **2**, 281–285 (2009).
26. Garcia, H. E. *et al.* *Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation* Vol. 3 (Government Printing Office, 2010).
27. Key, R. M. *et al.* A global ocean carbon climatology: results from Global Data Analysis Project (GLODAP). *Glob. Biogeochem. Cycles* **18**, GB4031 (2004).
28. Marcott, S. A. *et al.* Centennial-scale changes in the global carbon cycle during the last deglaciation. *Nature* **514**, 616–619 (2014).
29. Anderson, R. F. *et al.* Wind-driven upwelling in the Southern Ocean and the deglacial rise in atmospheric CO<sub>2</sub>. *Science* **323**, 1443–1448 (2009).
30. Taylor, S. R. & McLennan, S. M. *The Continental Crust: Its Composition and Evolution. An Examination of the Geochemical Record Preserved in Sedimentary Rocks* (Blackwell Scientific, 1985).
31. Jouzel, J. *et al.* Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science* **317**, 793–796 (2007).
32. Veres, D. *et al.* The Antarctic ice core chronology (AICC2012): an optimized multi-parameter and multi-site dating approach for the last 120 thousand years. *Clim. Past* **9**, 1733–1748 (2013).
33. Ahn, J. & Brook, E. J. Siple Dome ice reveals two modes of millennial CO<sub>2</sub> change during the last ice age. *Nature Commun.* **5**, 3723 (2014).
34. Bereiter, B. *et al.* Mode change of millennial CO<sub>2</sub> variability during the last glacial cycle associated with a bipolar marine carbon seesaw. *Proc. Natl Acad. Sci. USA* **109**, 9755–9760 (2012).

**Acknowledgements** S.L.J. and A.M.-G. were funded by the Swiss National Science Foundation (grants PP00P2-144811 and PZ00P2\_142424, respectively), E.D.G. by NSERC, and R.F.A. by the US NSF. Sediment samples were provided by the core repository at the Lamont-Doherty Earth Observatory. Computational resources were provided to E.D.G. by Compute Canada and the Canadian Foundation for Innovation. We thank C. Buizert, H. Fischer, F. Herman and T. Pedersen for discussions.

**Author Contributions** S.L.J. and R.F.A. conceived the study and S.L.J. wrote the first iteration of the manuscript. All co-authors provided input to the final version. S.L.J. oversaw the elemental analysis, while R.F.A. supervised the isotopic measurements. E.D.G. provided the climate model outputs and generated the statistical analysis. A.M.-G. refined the age model for core TN057-14PC.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.L.J. ([samuel.jaccard@geo.unibe.ch](mailto:samuel.jaccard@geo.unibe.ch)).

## METHODS

**Material and sediment composition.** TN057-13PC (53.2°S, 5.1°E, 2,848 m) and TN057-14PC (52.0°S, 4.5°E, 3,648 m) were retrieved from the Atlantic sector of the Southern Ocean, south of the present-day position of the Antarctic Polar Front (Fig. 1). The cores contain diatomaceous ooze sequences with discontinuous carbonate-bearing intervals and occasional ice-rafted debris layers. The deglacial sequence in core TN057-14PC is disturbed and thus was not further considered (Extended Data Fig. 6).

**Age model.** The TN057-13PC age model is taken, unchanged, from ref. 29. The stratigraphy of sediment core TN057-14PC has been updated on the basis of graphical correlation between the relative abundance of diatom *E. antarctica*<sup>29</sup>, a proxy for sea-surface temperature, and EDC deuterium ( $\delta D$ ) (ref. 31) expressed on the new AICC2012 age model<sup>32</sup>, assuming an in-phase relationship (Fig. 3). Sedimentation rates are high, typically in the range  $<15\text{--}60\text{ cm kyr}^{-1}$ , for glacial and peak interglacial intervals, respectively.

**Analytical methods.**  $^{230}\text{Th}$ -normalized fluxes were evaluated as described in ref. 35. Concentrations of U and Th isotopes were determined by isotope dilution inductively coupled plasma mass spectrometry (ICP-MS)<sup>36</sup>. The authigenic U content of each sample was evaluated using the measured  $^{238}\text{U}$  and  $^{232}\text{Th}$  concentrations using an activity ratio of 0.5 for the detrital endmember<sup>37</sup>. Measured  $^{230}\text{Th}$  concentrations were corrected for the detrital contribution using  $^{232}\text{Th}$  and for ingrowth produced by authigenic U, and then decay-corrected to the time of deposition to derive  $^{230}\text{Th}$ -normalized fluxes. Sedimentary aluminium (Al), manganese (Mn) and barium (Ba) concentrations were measured by inductively coupled plasma optical emission spectroscopy (ICP-OES; Varian Vista Pro) by ALS Minerals, Canada. Precision was better than 2% for replicate measurements. The biogenic fraction of barium—a useful proxy for integrated export production<sup>38</sup>—was determined using a detrital Ba/Al ratio of 0.0067 (ref. 30). The calculations of normative Ba/Al and Mn/Al—the excess Ba and Mn fractions relative to the detrital background, respectively—are based on the assumption that the composition of the Al-bearing phases of the terrigenous material remained constant in space and time.  $\text{CaCO}_3$  was quantified by coulometric  $\text{CO}_2$  determinations assuming no other carbonate-bearing phase was present. Precisions are  $\pm 3\%$  of the reported values.

**Redox-sensitive metals.** The sedimentary distribution of redox-sensitive trace metals can be influenced by a number of factors, including (1) the rate of organic matter delivery to the sea floor, (2) changes in sediment accumulation rate, and (3) changes in bottom-water oxygenation. Comparing the abundance of multiple trace metals (in our case, Mn and U) in sediments with biogenic flux to the sea floor reconstructed using  $^{230}\text{Th}$  normalization permits these processes to be distinguished. It follows that relative change of oxygen concentrations at the water–sediment interface can be inferred.

The sedimentary geochemistry of Mn is dominated by the redox control of its speciation, with higher oxidation states (Mn(III) and Mn(IV)) occurring as relatively insoluble oxyhydroxides in well oxygenated environments and the lower oxidation state (Mn(II)) being much more soluble in oxygen-depleted settings<sup>39</sup>. Mn enrichments in the form of Mn-oxide coatings are observed in deep sea sediments where oxic conditions prevail to greater sediment depths as a result of low organic matter respiration rates and well ventilated bottom waters. In sediments that are deposited under reducing conditions, there is no accumulation of Mn oxides and the Mn concentrations of buried deposits are low and entirely controlled by their insoluble detrital fraction. The presence of Mn concentration in excess of what can be expected from the detrital fraction thus suggests that the host sediment must have accumulated under oxygen-replete conditions.

Uranium behaves conservatively in oxygenated seawater and is present as U(VI). One of the primary removal mechanisms for U from seawater is via diffusion across the water–sediment interface of reducing sediments<sup>40</sup>. It appears that for sedimentary redox conditions near those for conversion of Fe(III) to Fe(II) (that is, suboxic conditions), soluble U(VI) is reduced to insoluble U(IV) (ref. 41). This lowers the pore-water U concentration, and dissolved U from the overlying bottom water can diffuse into the sediment along the declining concentration gradient, leading to an enrichment of authigenic (or excess) U in the sediment.

The emplacement of both Mn and U authigenic mineral phases are uncorrelated to  $^{230}\text{Th}$ -normalized opal fluxes (Figs 2 and 3, Extended Data Table 2). As a result, we infer that the variability in organic carbon delivery to the sea floor—and its subsequent respiration below the sediment–water interface—was not the primary factor controlling the sedimentary redox conditions.

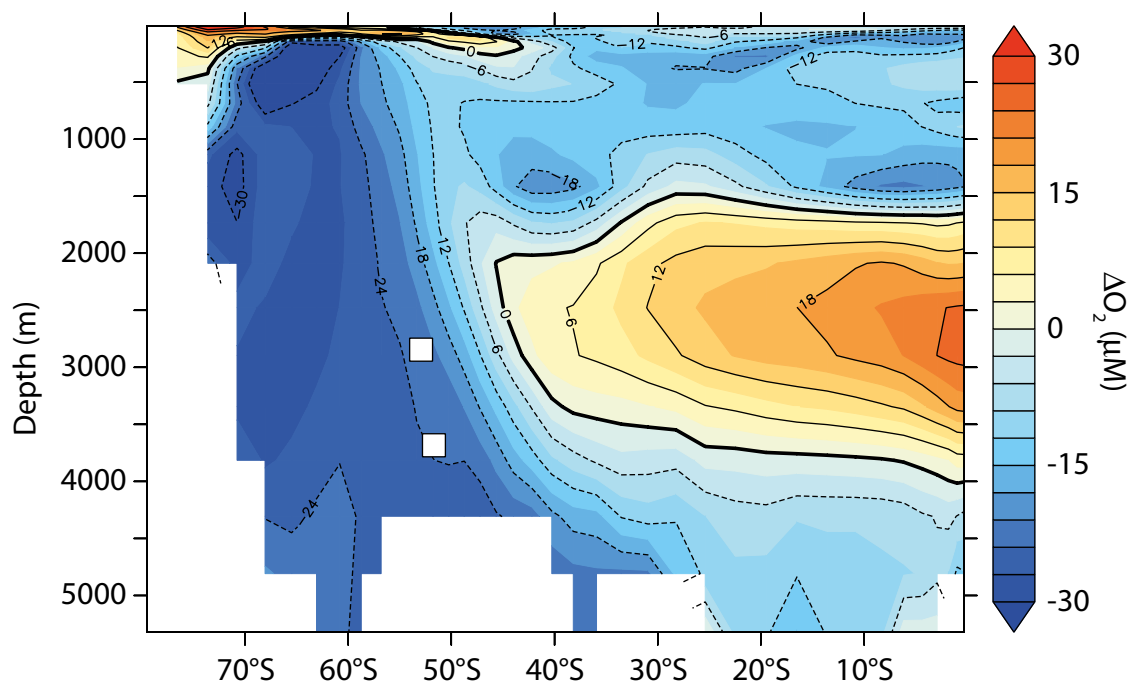
The sedimentation rates inferred for both sediment cores are high enough to minimize filtering by bioturbation and to prevent post-depositional remobilization of U upon deepening of the sedimentary oxicle. Manganese, on the other hand, is frequently remobilized to the sedimentary pore fluids under reducing conditions. Dissolved Mn can thus migrate in the sedimentary column and (re)precipitate when oxic conditions are encountered<sup>39</sup>. As such, large Mn enrichments

primarily reflect a shift from oxygen-depleted to oxygen-replete conditions. The sediment accumulation rate increased across the last glacial termination, potentially affecting trace-metal concentrations. However, both authigenic U and Mn/Al downcore records show opposite behaviours across the deglaciation, suggesting that sedimentary dilution was not the primary controlling factor driving the trace-metal distribution (Extended Data Fig. 7). Furthermore, the comparison between sedimentation rates and authigenic U concentrations in core TN057-14PC (Extended Data Fig. 5) clearly shows that the downcore authigenic U pattern was not dominated by changes in sediment accumulation. We thus posit that the observed authigenic enrichments of both U and Mn were sensitive to changing oxygen levels at the sediment–water interface.

**Model simulations.** The model simulations were carried out with the Geophysical Fluid Dynamics Laboratory (GFDL) coupled ocean–ice–atmosphere model CM2Mc, with the embedded biogeochemical model BLING<sup>42</sup>. Well equilibrated states of four simulations were used for the plots shown in Extended Data Fig. 1. Extended Data Fig. 1 shows the difference between two simulations with atmospheric  $\text{CO}_2$  concentrations of 220 parts per million and LGM ice sheet topography, but different obliquity (axial tilt). Strong convection and AABW production occurs under low obliquity (22°), while moderate convection occurs under high obliquity (24.5°). All other boundary conditions and parameters are identical in the two model runs. The plots were made from 100-year averages over the last century of 3,800-year simulations.

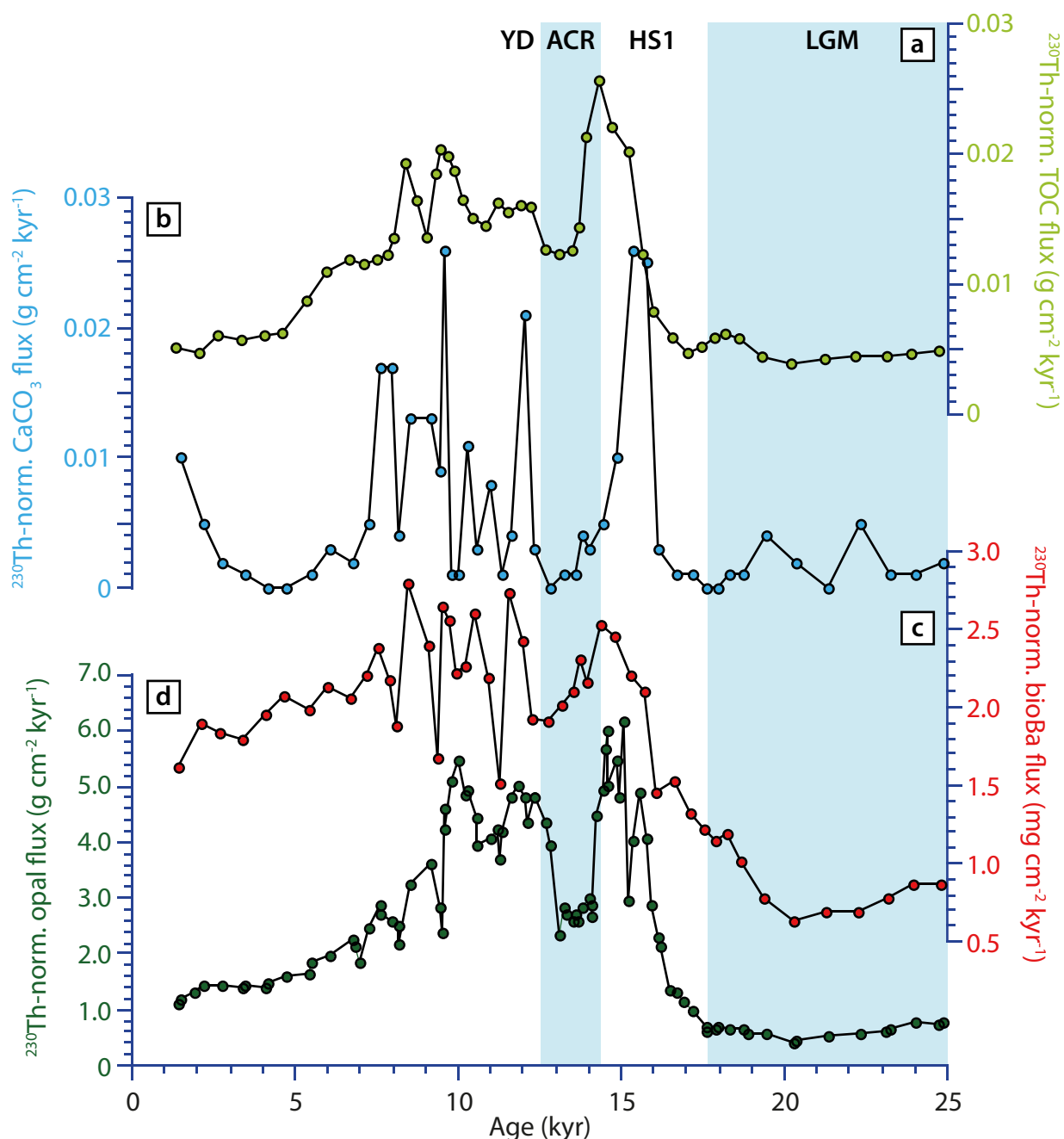
35. François, R., Frank, M., Rutgers van der Loeff, M. M. & Bacon, M. P.  $^{230}\text{Th}$  normalization: an essential tool for interpreting sedimentary fluxes during the late Quaternary. *Paleoceanography* **19**, PA1018 (2004).
36. Fleisher, M. Q. & Anderson, R. F. Assessing the collection efficiency of Ross Sea sediment traps using  $^{230}\text{Th}$  and  $^{231}\text{Pa}$ . *Deep Sea Res. Part II* **50**, 693–712 (2003).
37. Henderson, G. M. & Anderson, R. F. The U-series toolbox for paleoceanography. *Rev. Mineral. Geochem.* **52**, 493–531 (2003).
38. Dymond, J., Suess, E. & Lyle, M. Barium in deep-sea sediments: a geochemical proxy for paleoproductivity. *Paleoceanography* **7**, 163–181 (1992).
39. Calvert, S. E. & Pedersen, T. F. Sedimentary geochemistry of manganese: implication for the environment of formation of manganiferous black shales. *Econ. Geol.* **91**, 36–47 (1996).
40. Klinkhammer, G. P. & Palmer, M. R. Uranium in the oceans: where it goes and why. *Geochim. Cosmochim. Acta* **55**, 1799–1806 (1991).
41. Morford, J. L. & Emerson, S. The geochemistry of redox sensitive trace metals in sediments. *Geochim. Cosmochim. Acta* **63**, 1735–1750 (1999).
42. Galbraith, E. D., Gnanadesikan, A., Dunne, J. P. & Hiscock, M. R. Regional impacts of iron-light colimitation in a global biogeochemical model. *Biogeochemistry* **7**, 1043–1064 (2010).
43. Wagner, M. & Hendy, I. L. Trace metal evidence for a poorly ventilated glacial Southern Ocean. *Clim. Past Discuss.* **11**, 637–670 (2015).
44. Yu, J. *et al.* Deep South Atlantic carbonate chemistry and increased interoceanic deep water exchange during the last deglaciation. *Quat. Sci. Rev.* **90**, 80–89 (2014).
45. Lambert, F., Bigler, M., Steffensen, J. P., Hutterli, M. A. & Fischer, H. Centennial mineral dust variability in high-resolution ice core data from Dome C, Antarctica. *Clim. Past* **8**, 609–623 (2012).
46. Nielsen, S. H. H., Hodell, D. A., Kamenov, G., Guilderson, T. & Perfit, M. R. Origin and significance of ice-rafted detritus in the Atlantic sector of the Southern Ocean. *Geochim. Geophys. Res.* **8**, Q12005 (2007).
47. Kanfoush, S. L. *et al.* Millennial-scale instability of the Antarctic ice sheet during the last glacialization. *Science* **288**, 1815–1819 (2000).
48. Chase, Z., Anderson, R. F. & Fleisher, M. Q. Evidence from authigenic uranium for increased productivity of the glacial Subantarctic Ocean. *Paleoceanography* **16**, 468–478 (2001).
49. Anderson, R. F. *et al.* Late-Quaternary changes in productivity of the Southern Ocean. *J. Mar. Syst.* **17**, 497–514 (1998).
50. Kumar, N. *et al.* Increased biological productivity and export in the glacial Southern Ocean. *Nature* **378**, 675–680 (1995).
51. Martínez-García, A. *et al.* Links between iron supply, marine productivity, sea surface temperature, and  $\text{CO}_2$  over the last 1.1 Ma. *Paleoceanography* **24**, PA1207 (2009).
52. Dezileau, L., Bareille, G. & Reyss, J.-L. Enrichissement en uranium authigène dans les sédiments glaciaires de l’océan Austral. *C. R. Geosci.* **334**, 1039–1046 (2002).
53. Rosenthal, Y., Boyle, E. A., Labeyrie, L. D. & Oppo, D. Glacial enrichments of authigenic Cd and U in Subantarctic sediments: a climatic control on the elements’ oceanic budget? *Paleoceanography* **10**, 395–413 (1995).
54. Bareille, G. *et al.* Glacial-interglacial changes in the accumulation rates of major biogenic components in the Southern Indian Ocean sediments. *J. Mar. Syst.* **17**, 527–539 (1998).
55. Chase, Z., Anderson, R. F., Fleisher, M. Q. & Kubik, P. W. Accumulation of biogenic and lithogenic material in the Pacific sector of the Southern Ocean during the past 40,000 years. *Deep Sea Res. Part II* **50**, 799–832 (2003).
56. François, R., Bacon, M. P., Altabet, M. A. & Labeyrie, L. D. Glacial/interglacial changes in sediment rain rate in the SW Indian sector of Subantarctic waters as recorded by  $^{230}\text{Th}$ ,  $^{231}\text{Pa}$ , U and  $\delta^{15}\text{N}$ . *Paleoceanography* **8**, 611–629 (1993).





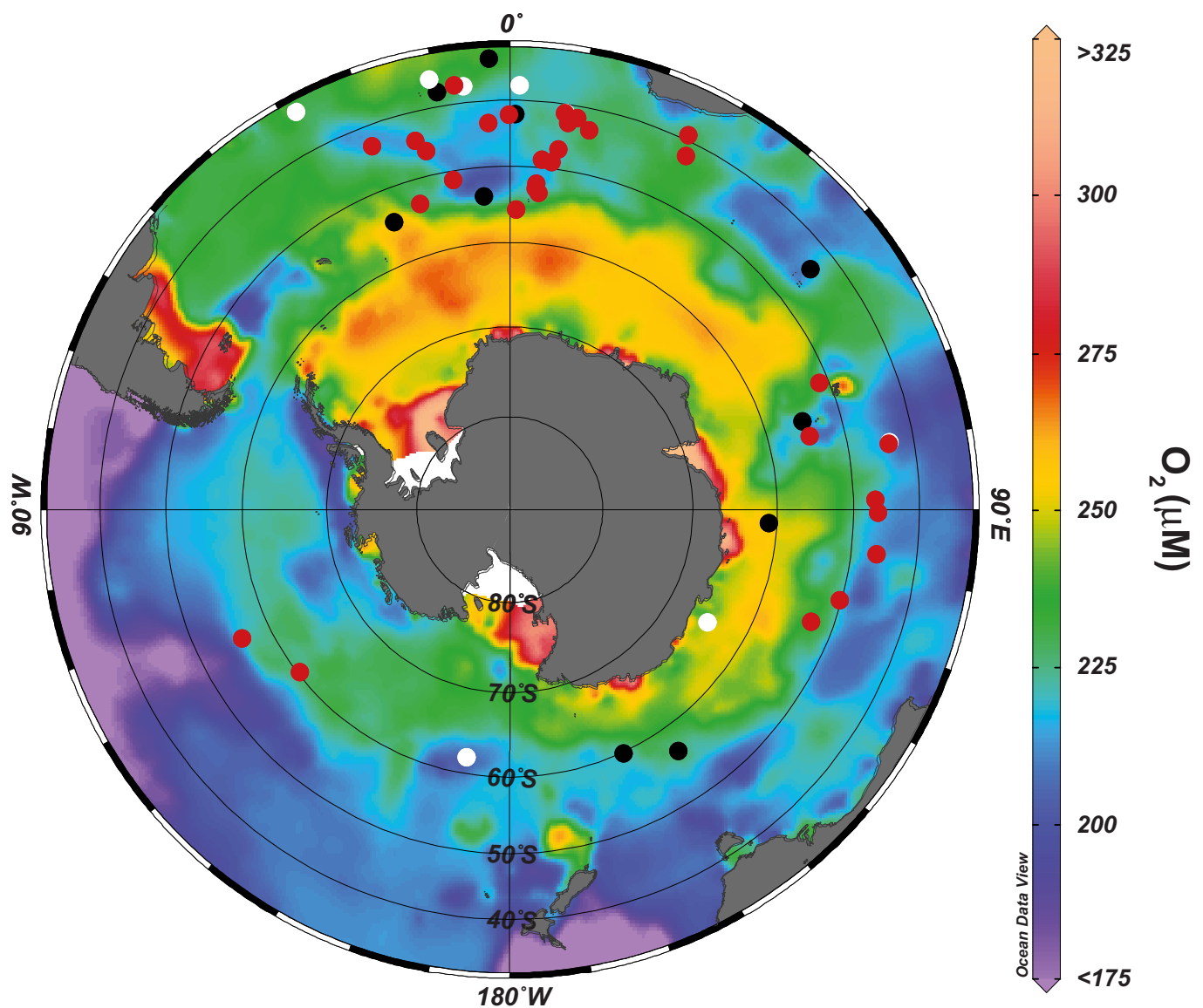
**Extended Data Figure 1 | Idealized model experiments illustrating the impact of AABW production on dissolved oxygen relative to the core locations.** Shaded contours show the difference in dissolved oxygen ( $\Delta O_2$ ) averaged between  $25^\circ$  W and  $10^\circ$  E, for a coupled model simulation

with strong Weddell convection compared to a simulation with moderate Weddell convection (Methods). Squares indicate the location of sediment cores TN057-13PC and TN057-14PC.



**Extended Data Figure 2 | Biogenic particle flux reconstructed by  $^{230}\text{Th}$  normalization for four independent proxies covering the last glacial termination at site TN057-13PC.** a,  $^{230}\text{Th}$ -normalized total organic carbon flux<sup>43</sup>. b,  $^{230}\text{Th}$ -normalized  $\text{CaCO}_3$  flux. c,  $^{230}\text{Th}$ -normalized biogenic barium (bioBa) flux. d,  $^{230}\text{Th}$ -normalized biogenic opal flux<sup>29</sup>.  $\text{CaCO}_3$  and bioBa data are from this study. The accumulation of biogenic

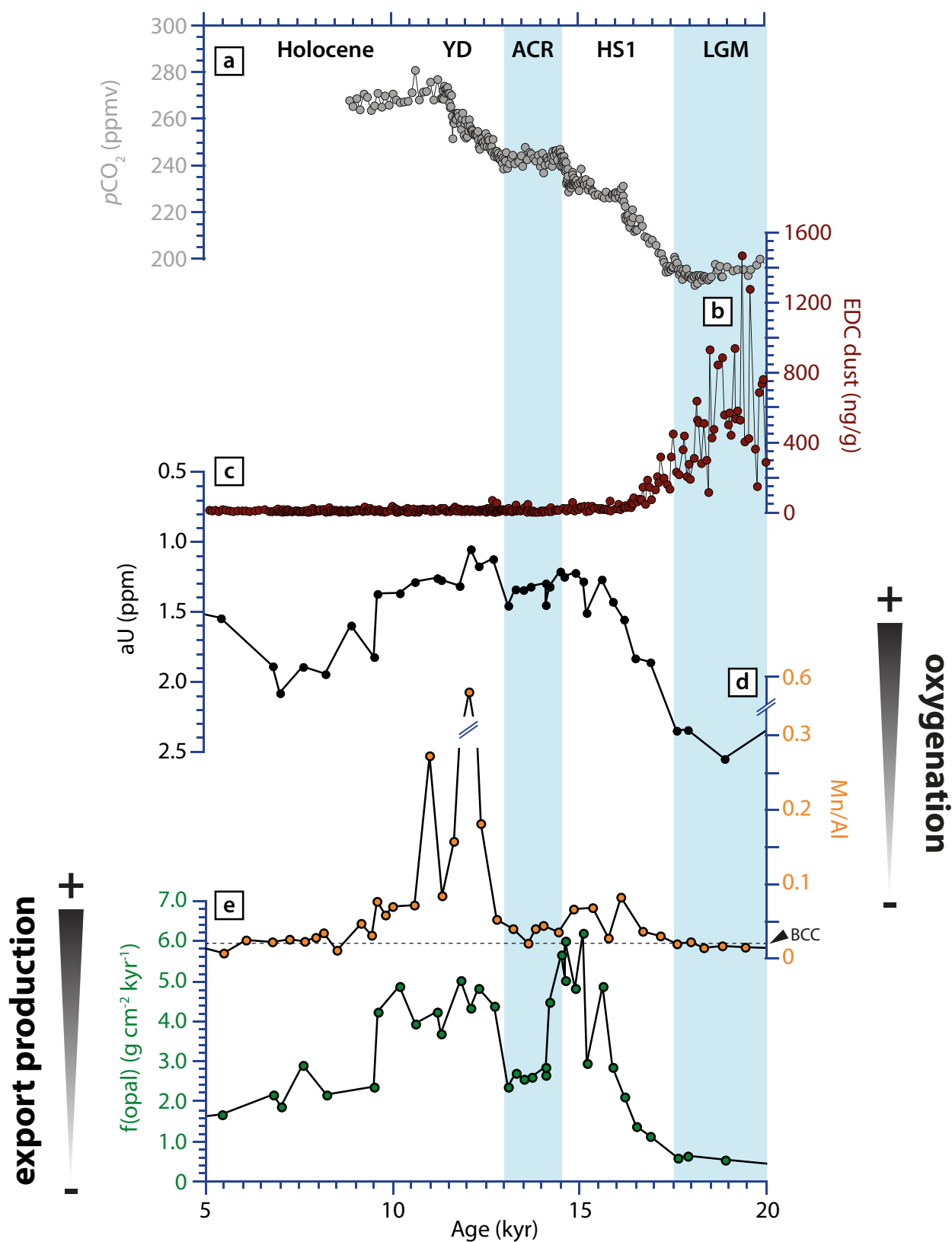
$\text{CaCO}_3$  above glacial background values during HS1 and the YD is consistent with enhanced ventilation of bottom waters during these intervals. Enhanced ventilation of bottom waters would have lowered the regenerated DIC concentration of the bottom water by releasing excess  $\text{CO}_2$  to the atmosphere, raising the  $[\text{CO}_3^{2-}]$  (ref. 44) and calcite saturation state of the bottom water<sup>9</sup> and thus reducing  $\text{CaCO}_3$  dissolution.



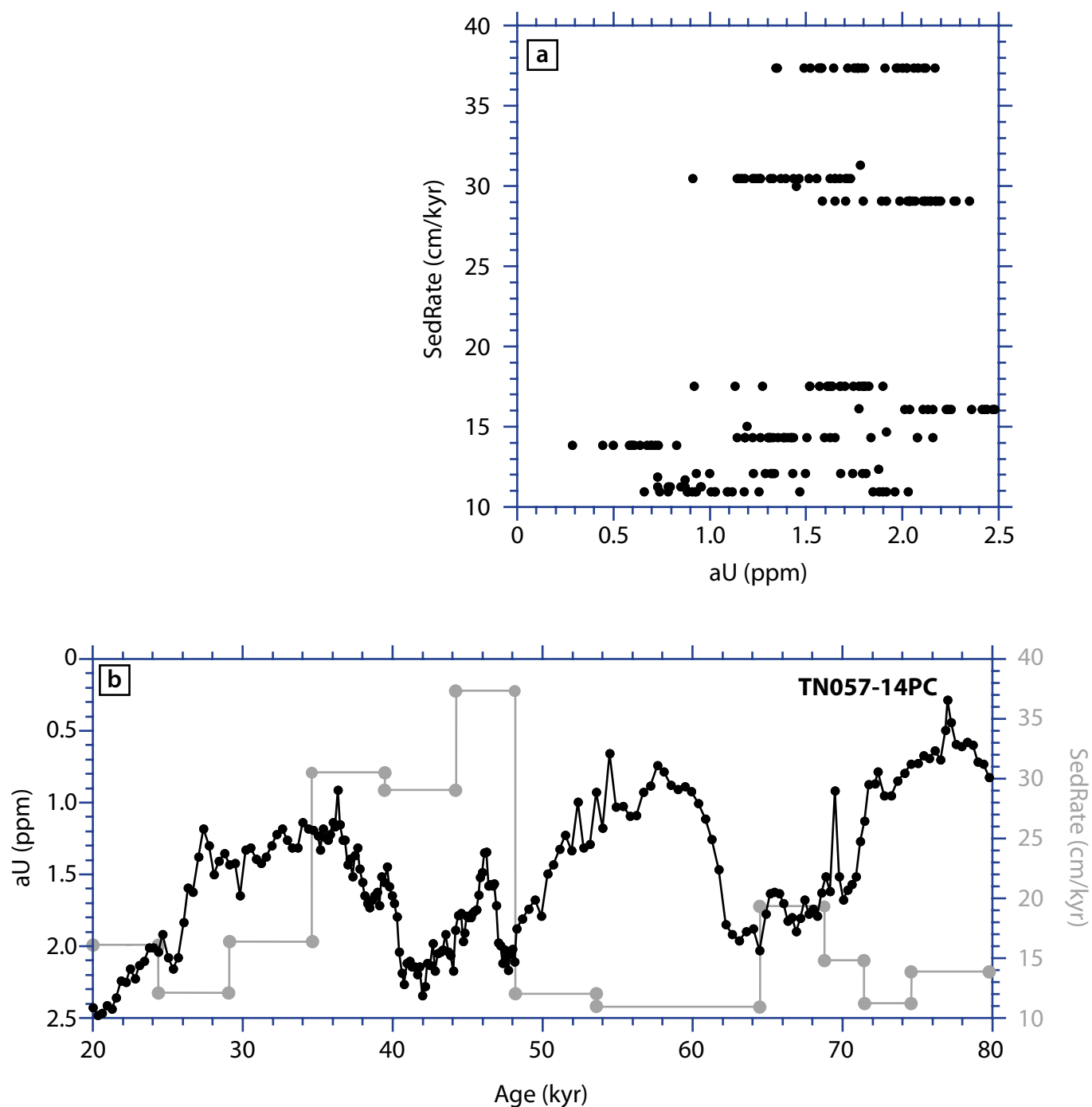
**Extended Data Figure 3 | Qualitative changes in oxygenation between the LGM and the Holocene.** Red/black dots indicate the location of sedimentary records for which authigenic U concentrations/mass accumulation rates were higher/lower during the LGM when compared to

the Holocene, respectively. White dots highlight cores where authigenic U concentrations did not change much between these two intervals (see Extended Data Table 1 for details). Shadings show the modern bottom water dissolved oxygen concentrations<sup>26</sup>.

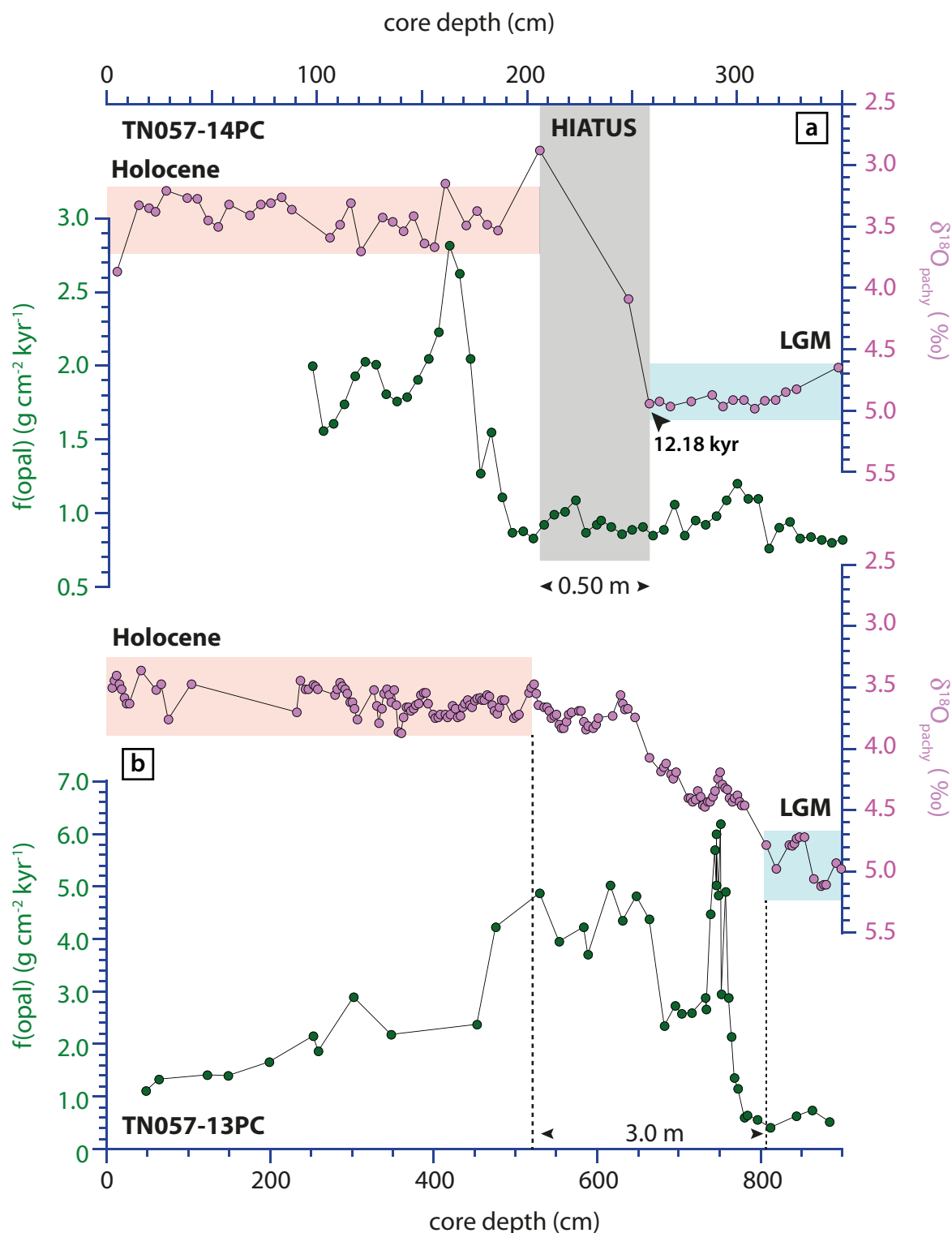




**Extended Data Figure 4 | Palaeoclimatic reconstructions covering the last glacial termination at site TN057-13PC compared with ice core records.** a, Atmospheric  $p\text{CO}_2$  (ref. 28). b, EPICA Dome C (EDC) dust flux<sup>45</sup>. c, Sedimentary authigenic U (aU) concentrations. d, Sedimentary Mn/Al. e,  $^{230}\text{Th}$ -normalized biogenic opal flux,  $f(\text{opal})$ <sup>29</sup>.



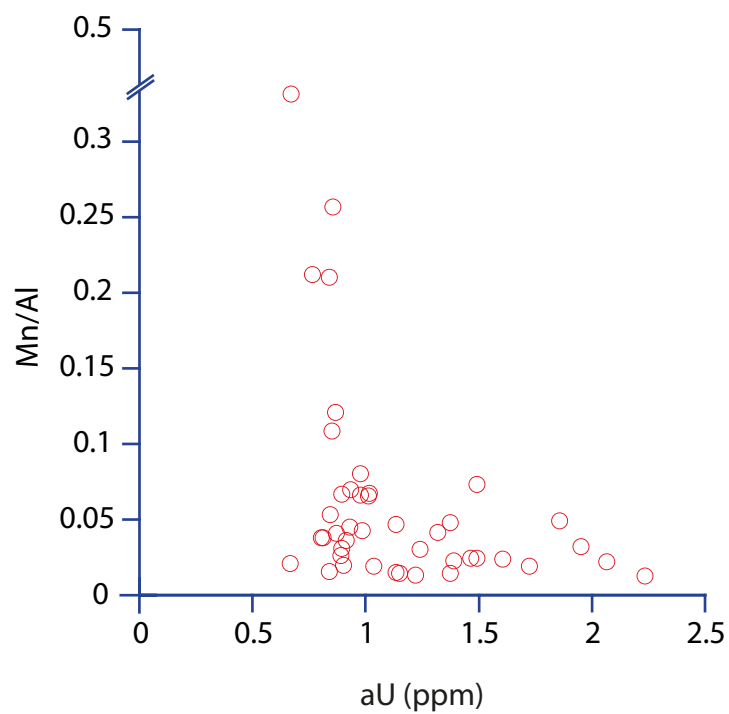
Extended Data Figure 5 | Comparison of bulk sediment accumulation rates and authigenic U concentrations in sediment core TN057-14PC for the interval 20–80 kyr ago. **a**, Bulk sediment accumulation rates (SedRate). **b**, Authigenic U concentrations.



**Extended Data Figure 6 | Comparison of the deglacial sequences at sites TN057-14PC and TN057-13PC.** **a**,  $\delta^{18}\text{O}_{\text{pachy}}$  (where 'pachy' refers to the planktonic foraminifera *Neogloboquadrina pachyderma*) (ref. 46) and  $^{230}\text{Th}$ -normalized biogenic opal flux<sup>28</sup> in core TN057-14PC. **b**,  $\delta^{18}\text{O}_{\text{pachy}}$  (ref. 47) and  $^{230}\text{Th}$ -normalized biogenic opal flux<sup>28</sup> in core TN057-13PC.

The grey shading highlights the disturbed portion of core TN057-14PC. The black triangle highlights the presence of planktonic foraminifera deposited during the Younger Dryas (that is, 12.18 kyr ago; ref. 46), which have been mixed down into late LGM sediments after the hiatus occurred.





Extended Data Figure 7 | Cross-plot of Mn/Al and authigenic U across the last glacial termination in sediment core TN057-13PC.

**Extended Data Table 1 | Available Southern Ocean low-resolution records permitting reconstruction of the authigenic U Holocene–LGM gradient compiled from the literature****Sedimentary records showing *higher* aU concentration/accumulation during the LGM (18–22kyr) than the Holocene (0–10kyr)**

Core	Latitude	Longitude	Depth (m)	References
VM27-196	-36.33	-7.57	1,106	Chase et al., 01 (ref. 48)
RC12-267	-38.68	25.45	4,144	Anderson et al., 98 (ref. 49)
TN057-21	-41.13	7.82	4,981	Kumar et al., 95 (ref. 50)
RC11-83	-41.60	9.72	4,718	Chase et al., 01 (ref. 48)
VM34-157	-41.95	26.42	3,636	Anderson et al., 98 (ref. 49)
VM22-109	-41.97	-0.25	733	Chase et al., 01 (ref. 48)
ODP1090	-42.60	8.50	3,700	Martinez-Garcia et al., 09 (ref. 51)
RC15-94	-42.90	-20.85	3,762	Kumar et al., 95 (ref. 50)
VM22-108	-43.18	-3.25	4,171	Kumar et al., 95 (ref. 50)
PS2082	-43.21	11.73	4,610	Frank et al., 00 (ref. 16)
RC11-120	-43.52	79.87	3,193	Anderson et al., 98 (ref. 49)
PS2498-1	-44.15	-14.49	3,783	Anderson et al., 14 (ref. 7)
MD88-770	-46.00	96.50	3,290	Dezileau et al., 02 (ref. 52)
MD88-769	-46.10	90.10	3,420	Rosenthal et al., 95 (ref. 53)
RC15-93	-46.10	-13.21	2,714	Anderson et al., 98 (ref. 49)
MD94-104	-46.50	88.10	3,460	Dezileau et al., 02 (ref. 52)
PS1754	-46.77	7.61	2,471	Frank et al., 00 (ref. 16)
RC13-254	-48.57	5.12	3,636	Kumar et al., 95 (ref. 50)
PS1756	-48.73	6.71	3,787	Frank et al., 00 (ref. 16)
MD88-772	-50.00	104.90	3,240	Bareille et al., 98 (ref. 54)
RC11-78	-50.86	-9.86	3,115	Chase et al., 01 (ref. 48)
MD80-304	-50.93	67.43	1,930	Rosenthal et al., 95 (ref. 53)
RC13-271	-51.98	4.51	3,634	Kumar et al., 95 (ref. 50)
PS1768	-52.60	4.48	3,270	Frank et al., 00 (ref. 16)
MD88-773	-52.90	109.87	2,460	Bareille et al., 98 (ref. 54)
RC11-77	-53.05	-16.45	4,098	Anderson et al., 98 (ref. 49)
TN057-13PC	-53.20	5.10	2,848	<i>this study</i>
MD84-552	-54.40	75.83	1,780	Dezileau et al., 02 (ref. 52)
PS1772	-55.46	1.17	4,135	Frank et al., 00 (ref. 16)
E11-2	-56.07	-115.09	3,111	Chase et al., 03 (ref. 55)
E20-10	-60.22	-127.05	4,474	Chase et al., 03 (ref. 55)

**Sedimentary records showing *lower* aU concentration/accumulation during the LGM than the Holocene**

VM24-221	-32.03	-2.82	4,204	Chase et al., 01 (ref. 48)
RC12-294	-37.26	-10.10	3,308	Chase et al., 01 (ref. 48)
TN057-20	-42.10	0.60	1,312	Chase et al., 01 (ref. 48)
MD84-527	-43.83	51.33	3,269	Francois et al., 93 (ref. 56)
RC13-259	-53.88	-4.93	2,677	Kumar et al., 95 (ref. 50)
RC11-76	-54.38	-22.13	5,229	Anderson et al., 98 (ref. 49)
E33-22	-54.93	-120.00	2,744	Chase et al., 03 (ref. 55)
MD84-551	-55.10	73.30	2,230	Bareille et al., 98 (ref. 54)
MD84-787	-56.40	145.30	3,020	Bareille et al., 98 (ref. 54)
E27-23	-59.62	155.24	3,215	Anderson et al., 09 (ref. 29)
KR8830	-61.00	93.20	4,300	Bareille et al., 98 (ref. 54)
E17-9	-63.08	-135.00	4,849	Chase et al., 03 (ref. 55)

**Sedimentary records showing no significant difference in aU concentration/accumulation between the LGM and the Holocene**

VM24-240	-31.73	-28.20	4,327	Chase et al., 01 (ref. 48)
VM24-229	-34.45	-10.60	4,202	Chase et al., 01 (ref. 48)
VM27-197	-36.68	-6.27	4,089	Chase et al., 01 (ref. 48)
RC13-243	-36.90	1.33	4,790	Chase et al., 01 (ref. 48)
VM14-65	-41.00	8.00	4,825	Chase et al., 01 (ref. 48)
MD94-102	-43.50	79.80	3,205	Bareille et al., 98 (ref. 54)
NBP9802-6	-61.88	-169.98	3,245	Chase et al., 03 (ref. 55)
NBP9802-5	-63.11	-169.74	2,940	Chase et al., 03 (ref. 55)
MD88-791	-64.70	119.50	3,150	Bareille et al., 98 (ref. 54)

Records from references 7, 16, 29 and 48–56.

Extended Data Table 2 | Statistical correlation between sedimentary proxies for the interval 20–82 kyr ago

	1090 f(Fe)	1090 f(alk)	TN057 aU	TN057 f(opal)	$p\text{CO}_2$
1090 f(Fe)					
1090 f(alk)	.62; 0.0000				
TN057 aU	.52; 0.0000	.64; 0.0000			
TN057 f(opal)	.24; 0.0000	.18; 0.0000	.13; 0.0000		
$p\text{CO}_2$	.71; 0.0000	.57; 0.0000	.58; 0.0000	.32; 0.0000	

<sup>230</sup>Th-normalized Fe and alkenone fluxes were measured at ODP site 1090 (ref. 6). Sedimentary U concentrations and <sup>230</sup>Th-normalized opal fluxes<sup>29</sup> were measured at TN057-14PC;  $p\text{CO}_2$  levels were measured on Antarctic ice<sup>33,34</sup>. Numbers in red represent the correlation coefficient,  $r^2$ ; the blue numbers indicate the  $P$  values, which are less than  $10^{-20}$  in all cases.



# Biomass resilience of Neotropical secondary forests

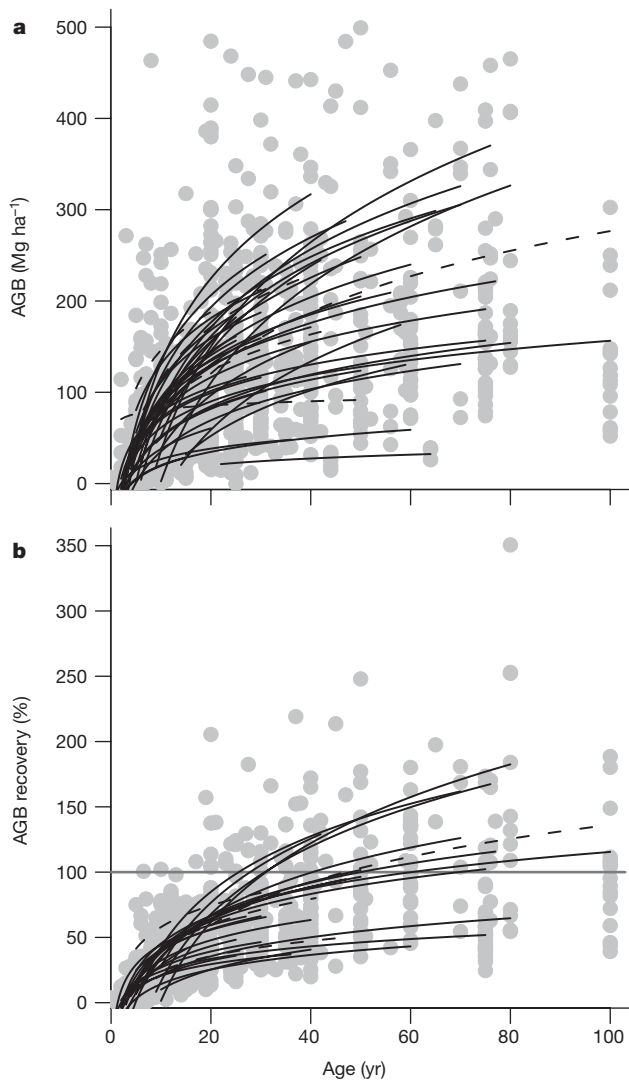
Lourens Poorter<sup>1</sup>, Frans Bongers<sup>1</sup>, T. Mitchell Aide<sup>2</sup>, Angélica M. Almeyda Zambrano<sup>3</sup>, Patricia Balvanera<sup>4</sup>, Justin M. Becknell<sup>5</sup>, Vanessa Boukili<sup>6</sup>, Pedro H. S. Brancalion<sup>7</sup>, Eben N. Broadbent<sup>3</sup>, Robin L. Chazdon<sup>6</sup>, Dylan Craven<sup>8,9,10</sup>, Jarcilene S. de Almeida-Cortez<sup>11</sup>, George A. L. Cabral<sup>11</sup>, Ben H. J. de Jong<sup>12</sup>, Julie S. Denslow<sup>13</sup>, Daisy H. Dent<sup>14,15</sup>, Saara J. DeWalt<sup>16</sup>, Juan M. Dupuy<sup>17</sup>, Sandra M. Durán<sup>18</sup>, Mario M. Espírito-Santo<sup>19</sup>, María C. Fandino<sup>20</sup>, Ricardo G. César<sup>7</sup>, Jefferson S. Hall<sup>8</sup>, José Luis Hernandez-Stefanoni<sup>17</sup>, Catarina C. Jakovac<sup>1,21</sup>, André B. Junqueira<sup>22,23,24</sup>, Deborah Kennard<sup>25</sup>, Susan G. Letcher<sup>26</sup>, Juan-Carlos Licona<sup>27</sup>, Madelon Lohbeck<sup>1,28</sup>, Erika Marín-Spiotta<sup>29</sup>, Miguel Martínez-Ramos<sup>4</sup>, Paulo Massoca<sup>21</sup>, Jorge A. Meave<sup>30</sup>, Rita Mesquita<sup>21</sup>, Francisco Mora<sup>4,30</sup>, Rodrigo Muñoz<sup>30</sup>, Robert Muscarella<sup>31,32</sup>, Yule R. F. Nunes<sup>19</sup>, Susana Ochoa-Gaona<sup>12</sup>, Alexandre A. de Oliveira<sup>33</sup>, Edith Orihuela-Belmonte<sup>12</sup>, Marielos Peña-Claros<sup>1</sup>, Eduardo A. Pérez-García<sup>30</sup>, Daniel Piotto<sup>34</sup>, Jennifer S. Powers<sup>35,36</sup>, Jorge Rodríguez-Velázquez<sup>4</sup>, I. Eunice Romero-Pérez<sup>30</sup>, Jorge Ruiz<sup>37,38</sup>, Juan G. Saldarriaga<sup>39</sup>, Arturo Sanchez-Azofeifa<sup>18</sup>, Naomi B. Schwartz<sup>31</sup>, Marc K. Steininger<sup>40</sup>, Nathan G. Swenson<sup>41</sup>, Marisol Toledo<sup>27</sup>, Maria Uriarte<sup>31</sup>, Michiel van Breugel<sup>8,42,43</sup>, Hans van der Wal<sup>44</sup>, Maria D. M. Veloso<sup>19</sup>, Hans F. M. Vester<sup>45,46</sup>, Alberto Vicentini<sup>21</sup>, Ima C. G. Vieira<sup>47</sup>, Tony Vizcarra Bentos<sup>21</sup>, G. Bruce Williamson<sup>21,48</sup> & Danaë M. A. Rozendaal<sup>1,6,49</sup>

**Land-use change occurs nowhere more rapidly than in the tropics, where the imbalance between deforestation and forest regrowth has large consequences for the global carbon cycle<sup>1</sup>. However, considerable uncertainty remains about the rate of biomass recovery in secondary forests, and how these rates are influenced by climate, landscape, and prior land use<sup>2–4</sup>. Here we analyse aboveground biomass recovery during secondary succession in 45 forest sites and about 1,500 forest plots covering the major environmental gradients in the Neotropics. The studied secondary forests are highly productive and resilient. Aboveground biomass recovery after 20 years was on average 122 megagrams per hectare (Mg ha<sup>-1</sup>), corresponding to a net carbon uptake of 3.05 Mg C ha<sup>-1</sup> yr<sup>-1</sup>, 11 times the uptake rate of old-growth forests. Aboveground biomass stocks took a median time of 66 years to recover to 90% of old-growth values. Aboveground biomass recovery after 20 years varied 11.3-fold (from 20 to 225 Mg ha<sup>-1</sup>) across sites, and this recovery increased with water availability (higher local rainfall and lower climatic water deficit). We present a biomass recovery map of Latin America, which illustrates geographical and climatic variation**

**in carbon sequestration potential during forest regrowth. The map will support policies to minimize forest loss in areas where biomass resilience is naturally low (such as seasonally dry forest regions) and promote forest regeneration and restoration in humid tropical lowland areas with high biomass resilience.**

Over half of the world's tropical forests are not old-growth, but naturally regenerating forests<sup>5</sup>. We focus here on secondary forests that regrow after nearly complete removal of forest cover for agricultural use<sup>4</sup>. Despite their crucial role in human-modified tropical landscapes, the rate at which these forests will recover, and the extent to which they can provide equivalent levels of ecosystem services as the forests they replaced, remains uncertain. Several studies suggest that, once a certain threshold of disturbance has been reached, tropical forests may collapse<sup>6</sup> or switch to an alternative stable state<sup>7</sup>. However, in general, secondary tropical forests have rapid rates of carbon sequestration, with potentially large consequences for the global carbon cycle<sup>3</sup>. The magnitude of the tropical carbon sink is relatively well known for old-growth forest<sup>8,9</sup>, but current global estimates on carbon uptake and storage of secondary forests are rough

<sup>1</sup>Forest Ecology and Forest Management Group, Wageningen University, PO Box 47, 6700 AA Wageningen, The Netherlands. <sup>2</sup>PO Box 23360, Department of Biology, University of Puerto Rico, San Juan, PR 00931-3360, Puerto Rico. <sup>3</sup>Spatial Ecology and Conservation Lab, Department of Geography, University of Alabama, Tuscaloosa, Alabama 35487, USA. <sup>4</sup>Instituto de Investigaciones en Ecosistemas y Sustentabilidad, Universidad Nacional Autónoma de México, CP58190, Morelia, Michoacán, México. <sup>5</sup>Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island 02912, USA. <sup>6</sup>Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut 06269, USA. <sup>7</sup>Department of Forest Sciences, Luiz de Queiroz College of Agriculture, University of São Paulo, Avenida Pádua Dias 11, 13418-900, Piracicaba, São Paulo, Brazil. <sup>8</sup>SI ForestGEO, Smithsonian Tropical Research Institute, Roosevelt Avenue, Tupper Building – 401, Balboa, Ancón, Panamá, Panamá. <sup>9</sup>German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany. <sup>10</sup>Institute for Biology, Leipzig University, Johannisallee 21, 04103 Leipzig, Germany. <sup>11</sup>Departamento de Botânica, Universidade Federal de Pernambuco, Pernambuco, CEP 50670-901, Brazil. <sup>12</sup>Department of Sustainability Science, El Colegio de la Frontera Sur, Unidad Campeche, Av. Rancho Polígono 2A, Parque Industrial Lerma, Campeche, Campeche, CP 24500, México. <sup>13</sup>Department of Ecology and Evolutionary Biology, Tulane University, New Orleans, Louisiana 70130, USA. <sup>14</sup>Smithsonian Tropical Research Institute, Roosevelt Avenue, Tupper Building – 401, Balboa, Ancón, Panamá, Panamá. <sup>15</sup>Biological and Environmental Sciences, University of Stirling, Stirling FK9 4LA, UK. <sup>16</sup>Department of Biological Sciences, Clemson University, 132 Long Hall, Clemson, South Carolina 29634, USA. <sup>17</sup>Centro de Investigación Científica de Yucatán, AC, Unidad de Recursos Naturales, Calle 43 No. 130, Colonia Chuburná de Hidalgo, CP 97200, Mérida, Yucatán, México. <sup>18</sup>Earth and Atmospheric Sciences Department, University of Alberta, Edmonton, Alberta T6G 2E3, Canada. <sup>19</sup>Departamento de Biología Geral, Universidade Estadual de Montes Claros, Montes Claros, Minas Gerais, CEP 39401-089, Brazil. <sup>20</sup>Fondo Patrimonio Natural para la Biodiversidad y Áreas Protegidas, Calle 72 No. 12-65 piso 6, Bogotá, Colombia. <sup>21</sup>Biological Dynamics of Forest Fragments Project, Environmental Dynamics Research Coordination, Instituto Nacional de Pesquisas da Amazonia, Manaus, Amazonas, CEP 69067-375, Brazil. <sup>22</sup>Centre for Crop Systems Analysis, Wageningen University, PO Box 430, 6700 AK Wageningen, The Netherlands. <sup>23</sup>Knowledge, Technology and Innovation Group, Wageningen University, PO Box 8130, 6700 EW Wageningen, The Netherlands. <sup>24</sup>Coordenação de Tecnologia e Inovação, Instituto Nacional de Pesquisas da Amazônia, Avenida André Araújo, 2936 – Aleixo, 69060-001 Manaus, Brazil. <sup>25</sup>Department of Physical and Environmental Sciences, Colorado Mesa University, 1100 North Avenue, Grand Junction, Colorado 81501, USA. <sup>26</sup>Department of Environmental Studies, Purchase College (State University of New York), Purchase, New York 10577, USA. <sup>27</sup>Instituto Boliviano de Investigación Forestal (IBIF), FCA-JAGRM, Casilla 6204, Santa Cruz de la Sierra, Bolivia. <sup>28</sup>World Agroforestry Centre (ICRAF), PO Box 30677 - 00100, Nairobi, Kenya. <sup>29</sup>Department of Geography, University of Wisconsin-Madison, 550 North Park Street, Madison, Wisconsin 53706, USA. <sup>30</sup>Departamento de Ecología y Recursos Naturales, Universidad Nacional Autónoma de México, México 04510 DF, México. <sup>31</sup>Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, New York 10027, USA. <sup>32</sup>Section of Ecoinformatics and Biodiversity, Department of Bioscience, Aarhus University, Aarhus 8000, Denmark. <sup>33</sup>Departamento de Ecología, Instituto de Biociências, Universidade de São Paulo, Rua do Matão, travessa 14, No. 321, São Paulo, CEP 05508-090, Brazil. <sup>34</sup>Universidade Federal do Sul da Bahia, Centro de Formação em Ciências Agroflorestais, Itabuna-BA, 45613-204, Brazil. <sup>35</sup>Department of Ecology, Evolution, & Behavior, University of Minnesota, Saint Paul, Minnesota 55108, USA. <sup>36</sup>Department of Plant Biology, University of Minnesota, Saint Paul, Minnesota 55108, USA. <sup>37</sup>School of Social Sciences, Geography Area, Universidad Pedagógica y Tecnológica de Colombia (UPTC), Tunja, Colombia. <sup>38</sup>Department of Geography, 4841 Ellison Hall, University of California, Santa Barbara, California 93106, USA. <sup>39</sup>Cr 5 No 14-05, PO Box 412, Cota, Cundinamarca, Colombia. <sup>40</sup>4007 18th St Northwest, Washington DC 20011, USA. <sup>41</sup>Department of Biology, University of Maryland, College Park, Maryland 20742, USA. <sup>42</sup>Yale-NUS College, 12 College Avenue West, Singapore 138610. <sup>43</sup>Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 11754. <sup>44</sup>Departamento de Agricultura, Sociedad y Ambiente, El Colegio de la Frontera Sur - Unidad Villahermosa, 86280 Centro, Tabasco, México. <sup>45</sup>Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, PO Box 94248, 1090 GE Amsterdam, The Netherlands. <sup>46</sup>Bonhoeffer College, Bruggertstraat 60, 7545 AX Enschede, The Netherlands. <sup>47</sup>Museu Paraense Emílio Goeldi, CP 399, CEP 66040-170, Belém, Brazil. <sup>48</sup>Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana 70803-1705, USA. <sup>49</sup>Department of Biology, University of Regina, 3737 Wascana Parkway, Regina, Saskatchewan S4S 0A2, Canada.

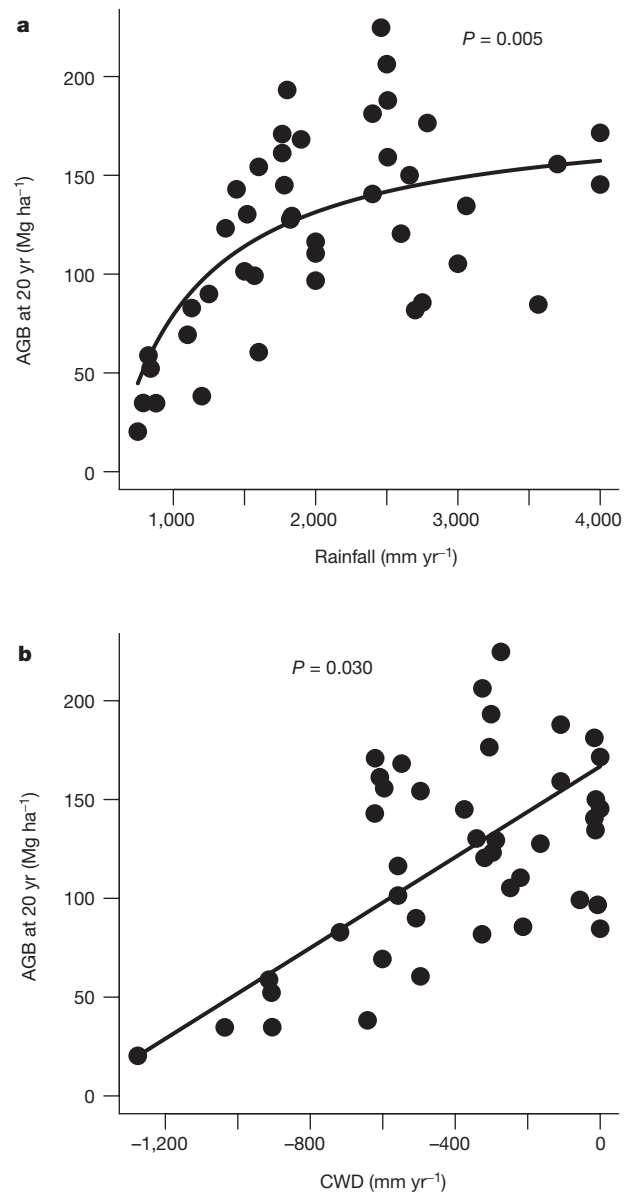


**Figure 1 | Relationship between forest biomass and stand age using chronosequence studies in Neotropical secondary forest sites.**

**a**, AGB ( $N = 44$ ); **b**, AGB recovery ( $N = 28$ ). Each line represents a different chronosequence. The original plots on which the regression lines are based are indicated in grey ( $N = 1,364$  for AGB,  $N = 995$  for AGB recovery). AGB recovery is defined as the AGB of the secondary forest plot compared with the median AGB of old-growth forest plots in the area, multiplied by 100. Significant relations (two-sided  $P \leq 0.05$ ) are indicated by continuous lines; non-significant relationships (two-sided  $P > 0.05$ ) are indicated by broken lines. Plots of 100 years old are also second-growth. See Extended Data Fig. 4 for the same figure with plots colour-coded by forest type.

or biased<sup>3,8,9</sup> and we urgently need better-founded biomass recovery rates and resilience estimates.

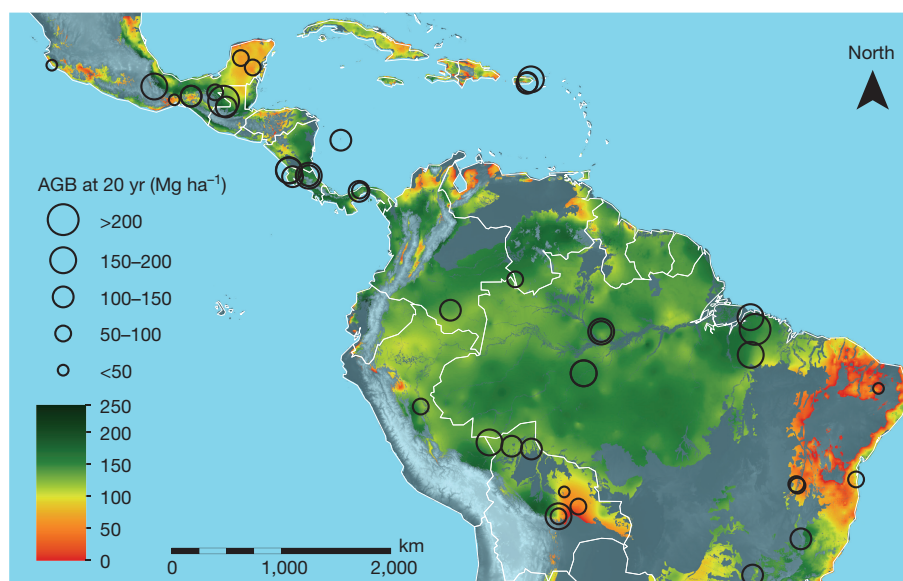
Secondary forests deliver a suite of ecosystem services that are closely linked to the biomass resilience of these forests<sup>10</sup>. We define biomass resilience as the ability to (1) recover quickly after disturbance (that is, the absolute recovery) and (2) return to its original (pre-deforestation) state (that is, the relative recovery). Few studies have quantified forest recovery directly by monitoring secondary forest plots over time, and they typically cover only a few years (but see ref. 11). Chronosequence studies have evaluated different plots at different times after abandonment<sup>12</sup> and shown that recovery rates vary dramatically with climate<sup>13</sup>, soils<sup>14</sup>, and management intensity<sup>15</sup>. However, the meta analyses so far<sup>13,16–18</sup> did not consider Neotropical-wide environmental gradients and were not based on original data, making direct comparisons impossible because of variation in methodology and allometric biomass equations used.



**Figure 2 | AGB after 20 years. a**, In relation to annual rainfall; **b**, in relation to CWD for Neotropical forest sites. Lines indicate predicted AGB at 20 years based on a multiple regression including  $1/\text{rainfall}$ , CWD, and rainfall seasonality ( $R^2 = 0.59$ ). Other variables were kept constant at the mean across sites (two-sided  $P < 0.005$  for  $1/\text{rainfall}$ ;  $P = 0.03$  for CWD). The third, less significant factor (rainfall seasonality) is shown in Extended Data Fig. 2.  $N = 43$  sites (one site was excluded because no climatic data were available).

Here we quantify rates of recovery of aboveground biomass (AGB) stocks in secondary forests, and assess to what extent recovery rates are driven by abiotic site conditions, landscape forest cover, and previous land use. We hypothesize that recovery rates increase with resource availability (high rainfall<sup>13</sup> and fertile soils) and decrease with degree of forest loss in the surrounding landscape matrix (implying lower seed availability), as well as with intensity of prior land use. We test these hypotheses using an unprecedented multi-site analysis including 45 sites, 1,468 plots, and  $>168,000$  stems, spanning the major environmental and latitudinal gradients in the Neotropics, and we present a biomass recovery map to support policy decisions for the Neotropics.

We compiled data from 45 chronosequence studies (Extended Data Table 1); we also included old-growth plots for 28 sites with no signs of recent human disturbance. In each plot all stems  $\geq 5$  cm diameter



**Figure 3 | Potential biomass recovery map of Neotropical secondary forests.** The total potential AGB accumulation over 20 years of lowland secondary forest growth was calculated on the basis of a regression equation relating AGB with annual rainfall ( $\text{AGB} = 135.17 - 103,950 \times 1/\text{rainfall} + 1.522 \times \text{rainfall seasonality} + 0.1148 \times \text{CWD}$ ; see Methods). The colour indicates the amount of forest cover recovery (red is low

recovery; green is high recovery). The 44 study sites are indicated by circles (the symbol on the ocean belongs to an island) and the size of the symbols scales with the AGB attained after 20 years. The grey areas do not belong to the tropical forest biome. The map focuses on lowland tropical forest (altitude  $<1,000$  m). See Extended Data Fig. 5 for a colour-blind-friendly map.

were measured and identified, their biomass estimated using allometric equations, and summed across all trees to obtain live plot AGB.

For each site, AGB was modelled as a function of time elapsed since abandonment: that is, since the main human activity (crop production, cattle ranching) ceased at the site. Biomass resilience was quantified in two ways. For absolute recovery rate, the site-specific models ( $N=44$ ; one outlier excluded) were used to predict AGB at 20 years, a representative age for secondary forests in the Neotropics. For relative recovery, AGB was expressed as a percentage of the median AGB of old-growth forest plots in the area. AGB recovery was related to climatic water availability, soil cation exchange capacity (CEC), forest cover, and prior land use using multiple regression analysis. The best-supported equation was used to produce a biomass recovery map of the Neotropics.

Recovery of AGB showed a saturating relationship with stand age for the individual sites. After 20 years of secondary succession, AGB accumulation varied 11.3-fold (from 20 to  $225 \text{ Mg ha}^{-1}$ , average  $121.8 \pm 7.5$  s.e.m.) across sites (Fig. 1a). Biomass recovery after 20 years, relative to old-growth forest AGB at the same site, varied from 25 to 85% across sites (Extended Data Fig. 1). Across sites, AGB after 20 years increased with annual precipitation ( $P < 0.001$ ), climatic water deficit (CWD;  $P < 0.001$ , a higher value here means a lower deficit), and rainfall seasonality ( $R^2 = 0.59$ ,  $P < 0.0001$ ; Fig. 2, Extended Data Fig. 2a and Extended Data Table 2). There was considerable variation at a given level of water availability, which is also reflected in the root mean squared error of  $31.7 \text{ Mg ha}^{-1}$ . CEC, forest cover, and prior land use were not significant predictors of AGB accumulation across sites (Extended Data Fig. 2). For relative recovery, not only rainfall ( $P = 0.040$ ) but also CEC ( $P = 0.027$ ) had a significant positive effect on biomass recovery (Extended Data Table 2 and Fig. 1). The biomass recovery map (Fig. 3, see Extended Data Fig. 3 for an uncertainty map) shows high variation in potential AGB accumulation rates across the Neotropics, with low rates in dry forests in Mexico and high rates in humid forests in southern Mexico, Costa Rica, and large parts of the Amazon.

We show that AGB of 20-year-old secondary forest varies more than 11-fold across sites, with potentially large consequences for the carbon storage and mitigation potential of tropical forests (Supplementary Information S1). The carbon sequestration and mitigation potential of secondary forests is immense. Average biomass after 20 years

corresponds to an annual net carbon uptake of  $3.05 \text{ Mg C ha}^{-1} \text{ yr}^{-1}$ , 11 times the uptake rates of Amazonian old-growth forests in 2010 ( $0.28 \text{ Mg C ha}^{-1} \text{ yr}^{-1}$  (ref. 19)) and 2.3 times the uptake rates of selectively logged Amazonian forests ( $1.33 \text{ Mg C ha}^{-1} \text{ yr}^{-1}$  (ref. 20)) where reduced impact logging techniques have been applied. Clearly, anthropogenic disturbances open up the canopy, enhance light availability, and set the system back to an earlier successional stage, leading to lower standing biomass but faster growth rates of the remaining forest stand. Although second-growth forests have substantially lower carbon stocks and biodiversity than the old-growth forests they replace, their carbon sequestering potential is high. Therefore, we urge halting deforestation and advocate actions that promote natural regeneration in deforested areas. Recognition of the significant carbon mitigation potential and other important services and values delivered by second-growth forests<sup>21</sup> can be an important motivating factor to reach national targets for forest restoration inspired by the Convention on Biological Diversity Aichi Targets, the Bonn Challenge, and the New York Declaration on Forests.

Secondary forests differ dramatically in their biomass resilience, driven mainly by variation in water availability (Fig. 2, cf. ref. 13), although other factors play a role as well. Both higher rainfall and lower CWD increase plant water availability and extend the growing season, thereby increasing biomass growth of trees and forest stands<sup>22</sup>. Climate change scenarios predict less and/or more variable rainfall for several regions in the tropics<sup>23</sup>, which may potentially hamper biomass recovery and forest resilience in these regions.

Soil fertility (CEC) had a significant positive effect on relative biomass recovery, but no effect on absolute recovery. Perhaps the effect of CEC was weak because for many sites CEC was obtained from a global database rather than measured *in situ*, because large differences in macroclimate may overrule more subtle differences in soil fertility, or because forest productivity is often limited by phosphorus<sup>24</sup> or nitrogen<sup>25</sup>, which we did not measure. Forest cover in the surrounding matrix and prior land use did not affect biomass recovery (Extended Data Figs 1 and 2), perhaps because these were estimated. Forest recovery rate may slow down in isolated secondary forest patches because of reduced plant colonization and survival, harsher environmental conditions, and frequent disturbances. Biomass recovery did not differ between abandoned pastures and agricultural fields, possibly because



of high within-category variation in land-use intensity (such as use of fire and time under land use and frequency of cultivation cycles), which may strongly affect forest recovery<sup>15–17</sup>.

Dry and moist forests differ substantially in old-growth forest biomass, and relative recovery may therefore be a better indicator of forest resilience than absolute recovery rates. It takes a median time of 66 years to recover to 90% of old-growth values. However, several mid-successional sites of 40–100 years old attained higher biomass than old-growth forest (Fig. 1b), which has also been found by modelling studies<sup>26</sup>. This is most probably because of a high abundance of old trees of long-lived pioneer species in mid-successional sites, which tend to be large and store large amounts of carbon<sup>27</sup> before they die.

Several studies suggest that, once a certain threshold of disturbance is reached, tropical forests may collapse<sup>6</sup> or switch to an alternative stable state<sup>7</sup>. Our study shows that forests can be resilient, and that their biomass resilience strongly depends on water availability. By mapping potential for biomass recovery across the Neotropics, attention can be focused on particular areas that should be conserved or treated with extra caution because they are more difficult to restore (slow recovery). The recovery map can also be used to identify areas with high carbon sequestration potential for United Nations collaborative initiative on Reducing Emissions from Deforestation and forest Degradation (REDD+) programmes, or where (assisted) natural regeneration, or restoration and afforestation activities, may have the highest chance of success<sup>28,29</sup> (high recovery). Such a spatially explicit, resource-based approach paves the road towards a more sustainable design and management of human-modified tropical landscapes.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 7 May; accepted 2 December 2015.**

**Published online 3 February 2016.**

- Hansen, M. C. *et al.* High-resolution global maps of 21st-century forest cover change. *Science* **342**, 850–853 (2013).
- IPCC. *2006 IPCC Guidelines for National Greenhouse Gas Inventories* (Institute for Global Environmental Strategies, 2006).
- Grace, J., Mitchard, E. & Gloor, E. Perturbations in the carbon budget of the tropics. *Glob. Change Biol.* **20**, 3238–3255 (2014).
- Chazdon, R. L. *Second Growth: The Promise of Tropical Forest Regeneration in an Age of Deforestation* Ch. 11 (Univ. Chicago Press, 2014).
- Food and Agriculture Organization of the United Nations. *Global Forest Resources Assessment*. FAO Forestry Paper 163 (FAO, 2010).
- Banks-Leite, C. *et al.* Using ecological thresholds to evaluate the costs and benefits of set-asides in a biodiversity hotspot. *Science* **345**, 1041–1045 (2014).
- Hirota, M., Holmgren, M., Van Nes, E. H. & Scheffer, M. Global resilience of tropical forest and savanna to critical transitions. *Science* **334**, 232–235 (2011).
- Pan, Y. *et al.* A large and persistent carbon sink in the world's forests. *Science* **333**, 988–993 (2011).
- Saatchi, S. S. *et al.* Benchmark map of forest carbon stocks in tropical regions across three continents. *Proc. Natl Acad. Sci. USA* **108**, 9899–9904 (2011).
- Lohbeck, M., Poorter, L., Martínez-Ramos, M. & Bongers, F. Biomass is the main driver of changes in ecosystem process rates during tropical forest succession. *Ecology* **96**, 1242–1252 (2015).
- Rozendaal, D. M. A. & Chazdon, R. L. Demographic drivers of tree biomass change during secondary succession in northeastern Costa Rica. *Ecol. Appl.* **25**, 506–516 (2015).
- Chazdon, R. L. *et al.* Rates of change in tree communities of secondary Neotropical forests following major disturbances. *Phil. Trans. R. Soc. B* **362**, 273–289 (2007).
- Becknell, J. M., Kucek, L. K. & Powers, J. S. Aboveground biomass in mature and secondary seasonally dry tropical forests: a literature review and global synthesis. *For. Ecol. Mgmt* **276**, 88–95 (2012).
- Becknell, J. M. & Powers, J. S. Stand age and soils as drivers of plant functional traits and aboveground biomass in secondary tropical dry forest. *Can. J. Forest Res.* **44**, 604–613 (2014).
- Jakovac, C. C., Peña-Claros, M., Kuyper, T. W. & Bongers, F. Loss of secondary-forest resilience by land-use intensification in the Amazon. *J. Ecol.* **103**, 67–77 (2015).
- Zarin, D. J. *et al.* Legacy of fire slows carbon accumulation in Amazonian forest regrowth. *Front. Ecol. Environ.* **3**, 365–369 (2005).
- Marín-Spiotta, E., Cusack, D. F., Ostertag, R. & Silver, W. L. in *Post-Agricultural Succession in the Neotropics* (ed. Myster, R. W.) 22–72 (Springer, 2008).
- Martin, P. A., Newton, A. C. & Bullock, J. M. Carbon pools recover more quickly than plant biodiversity in tropical secondary forests. *Proc. R. Soc. B* **280**, <http://dx.doi.org/10.1098/rspb.2013.2236> (2013).
- Brienen, R. J. W. *et al.* Long-term decline of the Amazon carbon sink. *Nature* **519**, 344–348 (2015).
- Rutishauser, E. *et al.* Rapid tree carbon stock recovery in managed Amazonian forests. *Curr. Biol.* **25**, R787–R788 (2015).
- Bongers, F., Chazdon, R., Poorter, L. & Peña-Claros, M. The potential of secondary forests. *Science* **348**, 642–643 (2015).
- Toledo, M. *et al.* Climate is a stronger driver of tree and forest growth rates than soil and disturbance. *J. Ecol.* **99**, 254–264 (2011).
- IPCC. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (IPCC, 2014).
- Quesada, C. A. *et al.* Basin-wide variations in Amazon forest structure and function are mediated by both soils and climate. *Biogeosciences* **9**, 2203–2246 (2012).
- Davidson, E. A. *et al.* Recuperation of nitrogen cycling in Amazonian forests following agricultural abandonment. *Nature* **447**, 995–998 (2007).
- Fischer, R., Armstrong, A., Shugart, H. H. & Huth, A. Simulating the impacts of reduced rainfall on carbon stocks and net ecosystem exchange in a tropical forest. *Environ. Model. Softw.* **52**, 200–206 (2014).
- Masero, J., Asner, G. P., Dent, D. H., DeWalt, S. J. & Denslow, J. S. Scale-dependence of aboveground carbon accumulation in secondary forests of Panama: a test of the intermediate peak hypothesis. *For. Ecol. Mgmt* **276**, 62–70 (2012).
- Chazdon, R. L. Beyond deforestation: restoring forests and ecosystem services on degraded lands. *Science* **320**, 1458–1460 (2008).
- Birch, J. C. *et al.* Cost-effectiveness of dryland forest restoration evaluated by spatial analysis of ecosystem services. *Proc. Natl Acad. Sci. USA* **107**, 21925–21930 (2010).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This paper is a product of the 2ndFOR collaborative research network on secondary forests. We thank the owners of the secondary forest sites for access to their forests, all the people who have established and measured the plots, and the institutions and funding agencies that supported them. We thank J. Zimmerman for the use of plot data, and the following agencies for financial support: Australian Department of Foreign Affairs and Trade-DFAT, CGIAR-FTA, CIFOR, Colciencias grant 1243-13-16640, Consejo Nacional de Ciencia y Tecnología (SEP-CONACYT 2009-129740 for ReSerBos, CONACYT 33851-B), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq: 563304/2010-3, 562955/2010-0, 574008/2008-0 and PQ 307422/2012-7), FOMIX-Yucatan (YUC-2008-C06-108863), ForestGEO, Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG CRA APQ-00001-11), Fundación Ecológica de Cuixmala, Heising-Simons Foundation, HSB, ICETEX, Instituto Internacional de Educação do Brasil-IEB, Instituto Nacional de Serviços Ambientais da Amazônia -Servamb-INPA, Inter-American Institute for Global Change (Tropi-Dr Network CRN3-025) via a grant from the US National Science Foundation (grant GEO-1128040), Motta Family Foundation, NASA Terrestrial Ecology Program, National Science Foundation (NSF-CNH-RCN grant 1313788 for Tropical Reforestation Network: Building a Socioecological Understanding of Tropical Reforestation (PARTNERS), NSF DEB-0129104, NSF BCS-1349952, NSF Career Grant DEB-1053237, NSF DEB 1050957, 0639393, 1147429, 0639114, and 1147434), NUFFIC, USAID (BOLFOR), Science without Borders Program (CAPES/CNPq) grant number 88881.064976/2014-01, The São Paulo Research Foundation (FAPESP) grant 2011/06782-5 and 2014/14503-7, Silicon Valley Foundation, Stichting Het Kronendak, Tropenbos Foundation, University of Connecticut Research Foundation, Wageningen University (INREF Terra Preta programme and FOREFRONT programme). This is publication number 683 in the Technical Series of the Biological Dynamics of Forest Fragments Project BDFFP-INPA-SI. This study was partly funded by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 283093; Role Of Biodiversity In climate change mitigation (ROBIN).

**Author Contributions** L.P., F.B. and D.R. conceived the idea and coordinated the data compilations, D.R. analysed the data, L.P., F.B., E.N.B. and R.C. contributed to analytical tools used in the analysis, E.N.B. and A.M.A.Z. made the map, L.P. wrote the paper, and all co-authors collected field data, discussed the results, gave suggestions for further analyses and commented on the manuscript.

**Author Information** Plot-level AGB data of 41 sites are available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.82vr4>, and for four sites they can be requested from L.P. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.P. (louis.poorter@wur.nl).

## METHODS

**Site description.** We compiled the largest data set of chronosequence studies in the lowland Neotropics so far, including 1,468 plots established in 45 forest sites (Extended Data Table 1), containing 168,210 tree stems. The sites were located in eight countries, covering the full latitudinal range in the tropics varying from 20°N in Mexico to 22°S in Brazil. We focused on lowland sites, generally below 1,000 m altitude. Annual rainfall varied fivefold across sites (from 750 to 4,000 mm yr<sup>-1</sup>), thus all major lowland forest types were included. Soil CEC varied 33-fold across sites (from 2 to 65 cmol(+) kg<sup>-1</sup>). Forest cover in the landscape ranged from 31 to 100% across sites.

**Site conditions.** We assessed whether biomass recovery rates were related to abiotic site conditions (rainfall, soil fertility), and land-use intensity (prior land use and percentage forest cover in the surrounding matrix). For each site, mean annual rainfall was obtained from the nearest climatological station. Biomass recovery should ultimately respond to plant water availability, which is a function of annual rainfall and its seasonal distribution, topography, soil depth, and the water holding capacity of the soil. Rainfall seasonality (expressed as a coefficient of variation of rainfall) was obtained from WorldClim<sup>30</sup> and CWD (in millimetres per year) from [http://chave.ups-tlse.fr/pantropical\\_allometry.htm](http://chave.ups-tlse.fr/pantropical_allometry.htm). CWD is the amount of water lost during dry months (defined as months where evapotranspiration exceeds rainfall), and is calculated as the total rainfall minus evapotranspiration during dry months. This number is by definition negative, and sites with CWD of 0 are not seasonally water stressed. CEC (in centimoles of positive charge per kilogram of soil) was used as an indicator of soil fertility of the site. Ideally, local soil data should be used, but these were only available for 12 sites. For the other sites, CEC data were obtained from the Harmonized World Soil Database (HWSD)<sup>31</sup>. For those 12 sites, the local CEC and HWSD data are indeed strongly positively correlated ( $r = 0.84$ ,  $P < 0.0006$ ,  $n = 12$ ), indicating that the HWSD ranks the sites well in terms of CEC. When local soil fertility data were available, the CEC of the old-growth forest was used, as this indicates potential site fertility, which is also included in the world soil database. The HWSD database did not contain data on soil nitrogen and phosphorus. Phosphorus is thought to limit plant growth in highly weathered and leached tropical soils and correlates well with the productivity of old-growth Amazonian forests<sup>24</sup>. Nitrogen may be more limiting during the first decades of secondary succession because a large part of vegetation nitrogen is volatilized when the slashed vegetation is burned<sup>25</sup>, while soil preparation, crop cultivation, and nutrient leaching may further deplete soil nutrients. Indeed, early successional forests have relatively many nitrogen-fixing trees. To describe the landscape matrix, we estimated per site, for all plots combined, the average cover in the following five land use classes in a 1 km radius around the plots: mature forest, secondary forest, plantations/agroforestry systems, agriculture/pasture, other. This estimate is based on the long-term field experience of the authors at the site where they worked. We then calculated per site the average percentage total forest cover (of mature and secondary forest) in a 1 km radius around the plots. We used total forest cover because forests improve microclimatic conditions at the landscape level and act as important sources of seeds and biotic agents (mammal and bird dispersal agents, mycorrhizae) that may help to speed up the rate of secondary succession. We classified the prior land use of each site as cattle ranching or shifting cultivation (or a combination of these two), which are the two major land-use types where forest regeneration occurs in the Neotropics. We acknowledge that different land use practices, such as fire frequency<sup>16</sup> and length of cultivation<sup>15</sup>, could affect biomass recovery rates of abandoned fields, but it is very difficult to obtain that information in a consistent and quantitative way across all sites.

**Plots.** Plot size varied from 0.01 to 1 ha. Most plots were rectangular, with a length varying from 20 to 200 m and width ranging from 1 to 50 m. The mean plot area across all ~1,468 plots was 947 m<sup>2</sup>, and the mean plot area across all 45 sites was 1,273 m<sup>2</sup>. On average 33 (5–291) plots were established per chronosequence. Chronosequence sites varied in the range of stand ages across the included secondary forest plots (Extended Data Table 1). For 28 out of the 45 sites, old-growth plots were included. Old-growth forests did not have signs of recent human disturbance, and were at least 100 years old. In most cases, the sampling design was the same as for secondary forest plots; however, for five of the sites, old-growth plots were slightly larger, for one site smaller. In each plot, all woody trees, shrubs, and palms  $\geq 5$  cm stem diameter at breast height (at 1.3 m height) were measured for their diameter and identified to species, with the exception of three sites for which only data for trees  $\geq 10$  cm diameter at breast height were available.

**Tree biomass.** We first evaluated how estimates of live AGB varied on the basis of the application of three frequently used sets of allometric equations. We calculated live AGB for each tree using equations from refs 32–34 (Supplementary Information S2). The equations in ref. 32 are based on stem

diameter only, and the equations in refs 33, 34 are based on a combination of stem diameter and wood density. Palms were treated in the same way as trees<sup>35</sup>, which may slightly overestimate palm biomass. Biomass was summed across all trees to obtain live plot AGB. The three different equations from refs 32–34 resulted in small differences in absolute plot AGB estimates for dry forests (average AGB after 20 years ranged from 68 to 85 Mg ha<sup>-1</sup> among the three equations), but yielded larger differences in AGB estimates for moist forest (precipitation 1,500–2,500 mm yr<sup>-1</sup>, AGB after 20 years ranged from 142 to 179 Mg ha<sup>-1</sup>; Extended Data Fig. 6). For 20-year-old forest, the earlier widely used equations<sup>32,33</sup> predicted about 21–26% more biomass than the more recently developed equation (Extended Data Fig. 6). These differences become even greater with increasing forest age, with large consequences for assessing the magnitude of carbon uptake and storage in tropical forests (see, for example, ref. 19), as moist forests contribute most to the current forested area in the Neotropics. For further analyses we used the equation in ref. 34 because it provides an allometric equation based on the largest harvested tree data set analysed so far (and includes and extends the equation of ref. 33), including many secondary forest species for trees from wet and dry forests, for trees of most tropical continents. It is ecologically very relevant and methodologically elegant, as it uses a correction factor  $E$ . This bioclimatic stress variable takes temperature variability, precipitation variability, and drought intensity into account, and therefore continuous variation in site aridity. The equation in ref. 34 also takes variation in wood density into account, which is an important source of plot AGB variation across the Neotropics<sup>36</sup>. For one site (Providencia Island) for which no  $E$  value was available, we estimated  $E$  from a regression equation relating  $E$  to annual rainfall.

Data about species-specific wood density (in grams per cubic centimetre) came from the local sites, or from the Neotropical data of a global wood density database<sup>37,38</sup> (<http://datadryad.org/handle/10255/dryad.235>), henceforth referred to as 'Dryad'. For all estimates of wood density, we used the data source (local or Dryad) that had the highest level of taxonomic resolution. In the case of the same level of taxonomic resolution, we used the local data source. If no wood density information was available at the species level, then genus- or family-level wood density values were used, as wood density is phylogenetically strongly conserved<sup>39</sup>. Across all sites except Surutu 1 and 2, Cayey, and Bolpebra, 94.0% of the trees of the secondary and old-growth plots had been identified to species (range 74.4–100%), 97.9% to genus (range 90.8–100%), and 98.9% to family (range 94.9–100%). For the other four sites, 36.0–55.6% of the trees had been identified to species, 61.6–86% to genus, and 61.6–86% to family.

**Plot structure.** Biomass was summed across all trees in a plot to obtain live plot AGB and expressed on a per hectare basis (AGB, in megagrams per hectare). For one site (San Carlos) the original tree inventory data were not available anymore, so for this site we used the plot biomass that the author had estimated using locally derived allometric equations. For plots with a nested design, we calculated AGB per hectare by extrapolating the biomass of trees in smaller size classes that were measured in a part of the plot only to the total area measured, thus assuming that forest structure within the plot did not vary. For four sites, only trees  $\geq 10$  cm diameter at breast height were inventoried, and for these sites a correction factor was used to obtain the estimated plot AGB when trees between 5 and 10 cm diameter at breast height would have been included. Plot AGB was therefore multiplied by 1.18 for dry forest and by 1.03 for moist forest (compare ref. 40). Some plots had extremely high AGB values ( $> 500$  Mg ha<sup>-1</sup>), because of the occurrence of a single large tree, in combination with a small plot size. These outliers (2.5% of the total number of plots) were excluded from the analysis. One site (Marques de Comillas) had extremely high average AGB ( $> 100$  Mg ha<sup>-1</sup>) 2 years after land abandonment, because of a combination of many large remnant forest trees being left during land use, and a small plot size. This site was excluded from the analysis. Our analysis focused on AGB only, which should be a good indicator of carbon recovery over time, as below-ground biomass in roots scales tightly with AGB of trees<sup>41</sup>, and soil organic matter (which includes necromass of dead fallen trees and roots) is rather constant with time after abandonment<sup>18</sup>.

**Statistical analyses.** No statistical methods were used to predetermine sample size. To evaluate how AGB (estimated with the equation from ref. 34) recovered over time, we modelled AGB in each site as a function of natural-logarithm-transformed age. We included an intercept, as in some cases there was AGB at time zero because remnant trees were left during forest clearance. Biomass resilience was quantified in two complementary ways: the absolute recovery rate and the relative recovery rate. For absolute recovery, the site-specific models ( $N = 44$  sites) were used to predict AGB at 10 and 20 years, as these ages were included in the age range of most sites, and are representative ages for secondary forests. To calculate annual carbon sequestration rates, we used the average AGB attained



after 20 years, multiplied it by 0.5 (which is the average carbon value in dry biomass), and divided it by 20. To evaluate how fast AGB of secondary forests recovers towards old-growth forest values, we calculated the relative AGB recovery as the AGB of the secondary forest plot compared with the median AGB of old-growth forest plots in the area (as a percentage). For each site, we estimated the recovery time as the time needed to recover to 90% of old-growth values. We used time to recover to 90%, rather than to recover 100% of old-growth values, as it indicates when the biomass is close to old-growth values. To attain the full 100% may take much more time, as we modelled biomass build-up to show an asymptotic relationship with time. This study aimed to assess the rates of recovery of tree biomass under conditions where regeneration is occurring naturally. Sometimes, succession can be arrested owing to high-intensity land use or seed limitation. We did not target sites where succession was arrested, so it is possible that our 'average' conditions could have overestimated rates of recovery when considering all potential areas. However, neither did we specifically select sites that were regenerating particularly fast. Given the high number of sites and plots across the Neotropics that cover the entire range of typical land-use types and intensities, we are confident that we provide realistic rates of biomass accumulation for dominant land-use types and intensities in the Neotropics.

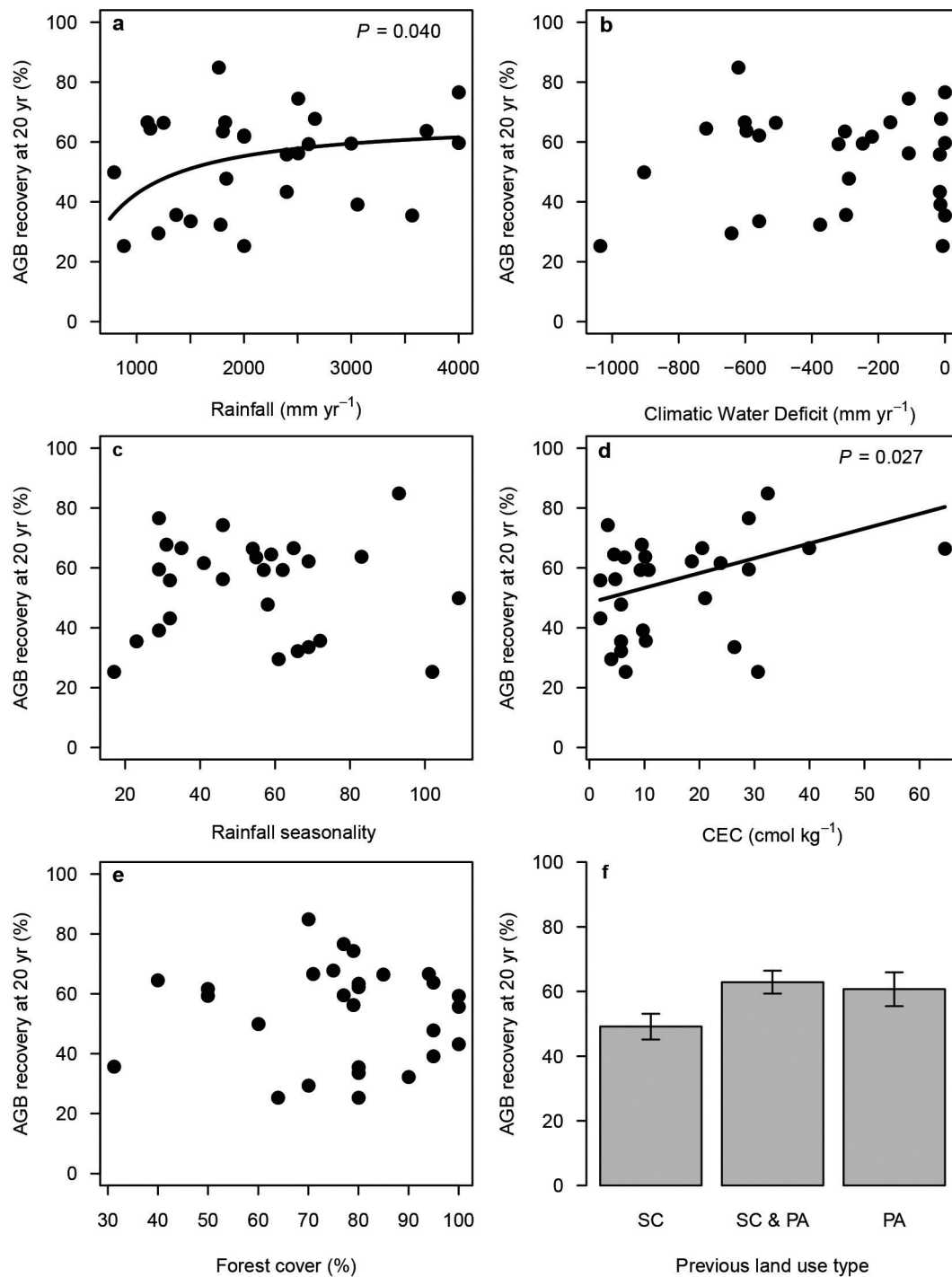
To evaluate how predicted absolute and relative recovery of AGB after 20 years varied with abiotic site conditions and land-use intensity, we used multiple regression analysis. We related AGB recovery to rainfall, rainfall seasonality, CWD, CEC, forest cover in the landscape matrix, and prior land use (pasture, mix of pasture and agriculture, agriculture) for 43 sites (for one island site no WorldClim data were available). AGB showed a saturating relationship with rainfall; therefore we used  $1/\text{rainfall}$  as a predictor rather than absolute rainfall or the natural logarithm of rainfall, which improved model fits. We compared models of all possible combinations of predictors based on Akaike's information criterion adjusted for small sample sizes ( $AIC_c$ ), and selected the best-supported model with the lowest  $AIC_c$ . All analyses were performed in R 3.1.2 (ref. 42).

On the basis of our regression analysis, we constructed a potential biomass recovery map of the Neotropics at a resolution of  $1 \text{ km}^2$ . First, a map of world ecoregions was obtained from the Nature Conservancy ([http://maps.tnc.org/gis\\_data.html](http://maps.tnc.org/gis_data.html); last accessed on 26 January 2015). Eight hundred and sixty-seven distinct units were categorized into 14 biomes in 8 biogeographical realms<sup>43</sup>. Of these 14 biomes, three are relevant to our results: (1) tropical and subtropical moist broadleaf forests, (2) tropical and subtropical dry broadleaf forests, and (3) tropical and subtropical coniferous forests. Our scaled-up study region was then defined as the full spatial extent encompassing these three biomes in Central or South America, the region in which our network of study sites was located. Total annual precipitation was calculated by summing the individual monthly totals provided by WorldClim. Data for mean annual rainfall (defined as the average of 1950–2000) and rainfall seasonality (defined as variable 'BIO15' in WorldClim) was obtained at a 30 s resolution (approximately  $1 \text{ km} \times 1 \text{ km}$ ) from WorldClim (<http://www.worldclim.org/current>), and CWD was obtained from [http://chave.ups-tlse.fr/pantropical\\_allometry.htm](http://chave.ups-tlse.fr/pantropical_allometry.htm). We then calculated the total potential AGB accumulation over 20 years of secondary forest growth (assuming that the initial year 0 condition was a fully cleared area), based on annual rainfall, rainfall seasonality, and CWD, because they were included in the best model explaining absolute recovery (Extended Data Table 2). The regression equation obtained in this study, relating AGB after 20 years with the climatic variables, was  $\text{AGB at 20 yr} = 135.17 - 103,950 \times 1/\text{rainfall} + 1.521983 \times \text{rainfall seasonality} + 0.1148 \times \text{CWD}$ . Estimated AGB would then indicate the absolute biomass recovery potential. We stress that this is a potential biomass recovery map based on the relationship between recovery and rainfall for our sites; realized local recovery may vary because of differences in local soil conditions, land use history, the surrounding matrix, and availability of seed sources. We also made an uncertainty map of potential biomass based on the confidence interval of the predicted mean AGB at 20 years obtained from the multiple regression. We calculated the confidence interval of the predicted AGB at 20 years for each pixel based on  $1/\text{rainfall}$ , CWD, and rainfall seasonality. The relative uncertainty for each pixel was calculated as  $100 \times (0.5 \times 95\% \text{ confidence interval of the mean})/\text{predicted AGB}$ . For the maps we focused on lowland forest ( $<1,000 \text{ m}$  altitude), between the Tropics of Cancer and Capricorn. To make the maps, we had to extrapolate very little beyond our climatic parameter space; our sites cover a rainfall range from  $750$  to  $4,000 \text{ mm yr}^{-1}$ . Sites with rainfall  $>4,000 \text{ mm yr}^{-1}$  (where we extrapolate with our equation) cover only 0.88% of the lowland Neotropics.

30. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
31. Nachtergaele, F., van Velthuisen, H., Verelst, L. & Wiberg, D. *Harmonized World Soil Database Version 1.2* (FAO and IIASA, 2012).
32. Pearson, R., Walker, S. & Brown, S. *Source Book for Land Use, Land-Use Change and Forestry Projects* (World Bank, 2005).
33. Chave, J. *et al.* Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia* **145**, 87–99 (2005).
34. Chave, J. *et al.* Improved allometric models to estimate the aboveground biomass of tropical trees. *Glob. Change Biol.* **20**, 3177–3190 (2014).
35. Clark, D. B. & Clark, D. A. Landscape-scale variation in forest structure and biomass in a tropical rain forest. *For. Ecol. Mgmt* **137**, 185–198 (2000).
36. Mitchard, E. T. A. *et al.* Markedly divergent estimates of Amazon forest carbon density from ground plots and satellites. *Glob. Ecol. Biogeogr.* **23**, 935–946 (2014).
37. Chave, J. *et al.* Towards a worldwide wood economics spectrum. *Ecol. Lett.* **12**, 351–366 (2009).
38. Zanne, A. E. *et al.* Data from: Towards a worldwide wood economics spectrum. <http://dx.doi.org/10.5061/dryad.234> (Dryad Digital Repository, 2009).
39. Chave, J. *et al.* Regional and phylogenetic variation of wood density across 2456 Neotropical tree species. *Ecol. Appl.* **16**, 2356–2367 (2006).
40. Poorter, L. *et al.* Diversity enhances carbon storage in tropical forests. *Glob. Ecol. Biogeogr.* **24**, 1314–1328 (2015).
41. Poorter, H. *et al.* How does biomass distribution change with size and differ among species? An analysis for 1200 plant species from five continents. *New Phytol.* **208**, 736–749 (2015).
42. R Core Team. R: a language and environment for statistical computing (R Foundation for Statistical Computing, 2014).
43. Olson, D. M. *et al.* Terrestrial ecoregions of the world: a new map of life on Earth. *Bioscience* **51**, 933–938 (2001).
44. Broadbent, E. N. *et al.* Integrating stand and soil properties to understand foliar nutrient dynamics during forest succession following slash-and-burn agriculture in the Bolivian Amazon. *PLoS ONE* **9**, e86042 (2014).
45. Peña-Claros, M. Changes in forest structure and species composition during secondary forest succession in the Bolivian Amazon. *Biotropica* **35**, 450–461 (2003).
46. Toledo, M. & Salick, J. Secondary succession and indigenous management in semideciduous forest fallows of the Amazon basin. *Biotropica* **38**, 161–170 (2006).
47. Kennard, D. K. Secondary forest succession in a tropical dry forest: patterns of development across a 50-year chronosequence in lowland Bolivia. *J. Trop. Ecol.* **18**, 53–66 (2002).
48. Steininger, M. K. Secondary forest structure and biomass following short and extended land use in central and southern Amazonia. *J. Trop. Ecol.* **16**, 689–708 (2000).
49. Piotto, D. *Spatial Dynamics of Forest Recovery after Swidden Cultivation in the Atlantic Forest of Southern Bahia, Brazil*. PhD thesis, Yale Univ. (2011).
50. Vieira, I. C. G. *et al.* Classifying successional forests using Landsat spectral properties and ecological characteristics in eastern Amazonia. *Remote Sens. Environ.* **87**, 470–481 (2003).
51. Williamson, G. B., Bentos, T. V., Longworth, J. B. & Mesquita, R. C. G. Convergence and divergence in alternative successional pathways in Central Amazonia. *Plant Ecol. Divers.* **7**, 341–348 (2014).
52. Madeira, B. G. *et al.* Changes in tree and liana communities along a successional gradient in a tropical dry forest in south-eastern Brazil. *Plant Ecol.* **201**, 291–304 (2009).
53. Cabral, G. A. L., Sampaio, E. V. S. B. & de Almeida-Cortez, J. S. Estrutura espacial e biomassa da parte aérea em diferentes estádios sucessionais de caatinga, em Santa Terezinha, Paraíba. *Rev. Bras. Geogr. Fis.* **6**, 566–574 (2013).
54. Junqueira, A. B., Shepard, G. H. & Clement, C. R. Secondary forests on anthropogenic soils conserve agrobiodiversity. *Biodivers. Conserv.* **19**, 1933–1961 (2010).
55. Vester, H. F. M. & Cleef, A. M. Tree architecture and secondary tropical rain forest development - a case study in Araracuara, Colombian Amazonia. *Flora* **193**, 75–97 (1998).
56. Ruiz, J., Fandino, M. C. & Chazdon, R. L. Vegetation structure, composition, and species richness across a 56-year chronosequence of dry tropical forest on Providencia Island, Colombia. *Biotropica* **37**, 520–530 (2005).
57. Saldarriaga, J. G., West, D. C., Tharp, M. L. & Uhl, C. Long-term chronosequence of forest succession in the upper Rio Negro of Colombia and Venezuela. *J. Ecol.* **76**, 938–958 (1988).
58. Powers, J. S., Becknell, J. M., Irving, J. & Perez-Aviles, D. Diversity and structure of regenerating tropical dry forests in Costa Rica: geographic patterns and environmental drivers. *For. Ecol. Mgmt* **258**, 959–970 (2009).
59. Chazdon, R. L., Brenes, A. R. & Alvarado, B. V. Effects of climate and stand age on annual tree dynamics in tropical second-growth rain forests. *Ecology* **86**, 1808–1815 (2005).
60. Letcher, S. G. & Chazdon, R. L. Rapid recovery of biomass, species richness, and species composition in a forest chronosequence in northeastern Costa Rica. *Biotropica* **41**, 608–617 (2009).
61. van Breugel, M., Martínez-Ramos, M. & Bongers, F. Community dynamics during early secondary succession in Mexican tropical rain forests. *J. Trop. Ecol.* **22**, 663–674 (2006).

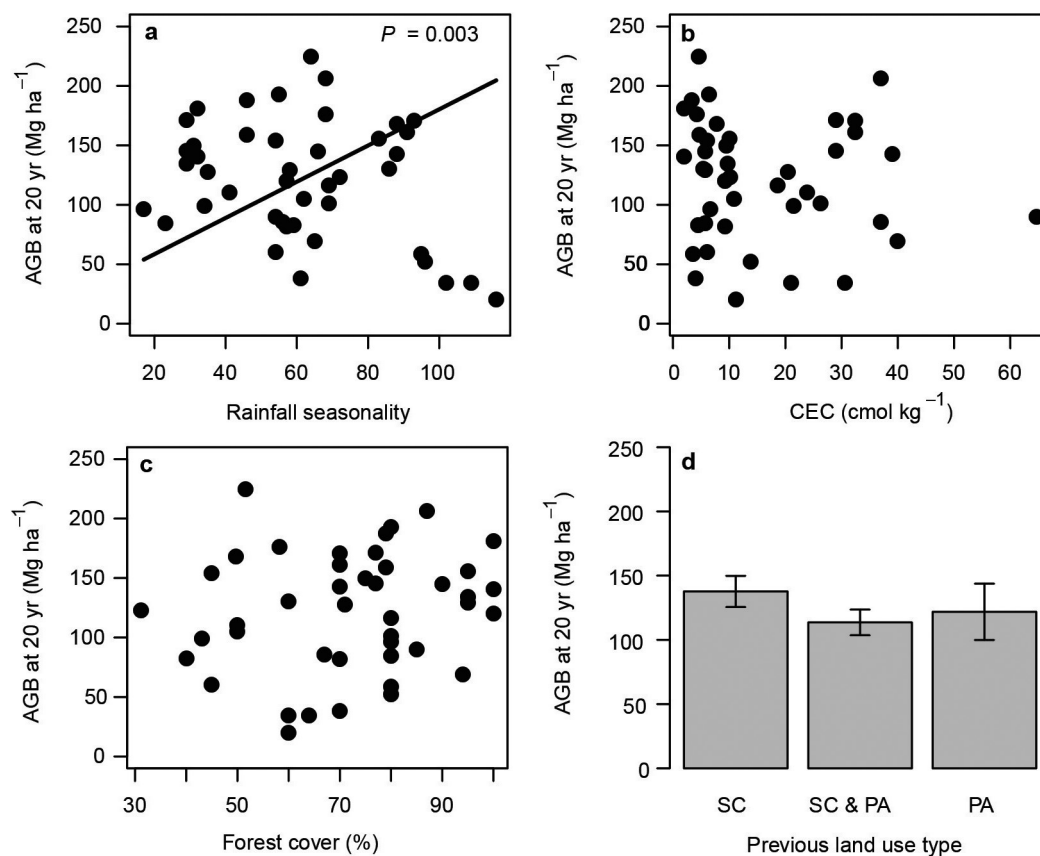


62. Mora, F. *et al.* Testing chronosequences through dynamic approaches: time and site effects on tropical dry forest succession. *Biotropica* **47**, 38–48 (2015).
63. Orihuela-Belmonte, D. E. *et al.* Carbon stocks and accumulation rates in tropical secondary forests at the scale of community, landscape and forest type. *Agric. Ecosyst. Environ.* **171**, 72–84 (2013).
64. Lebrija-Trejos, E., Bongers, F., Pérez-García, E. A. & Meave, J. A. Successional change and resilience of a very dry tropical deciduous forest following shifting agriculture. *Biotropica* **40**, 422–431 (2008).
65. Dupuy, J. M. *et al.* Patterns and correlates of tropical dry forest structure and composition in a highly replicated chronosequence in Yucatan, Mexico. *Biotropica* **44**, 151–162 (2012).
66. van Breugel, M. *et al.* Succession of ephemeral secondary forests and their limited role for the conservation of floristic diversity in a human-modified tropical landscape. *PLoS ONE* **8**, e82433 (2013).
67. Denslow, J. S. & Guzman, S. Variation in stand structure, light and seedling abundance across a tropical moist forest chronosequence, Panama. *J. Veg. Sci.* **11**, 201–212 (2000).
68. Marín-Spiotta, E., Ostertag, R. & Silver, W. L. Long-term patterns in tropical reforestation: plant community composition and aboveground biomass accumulation. *Ecol. Appl.* **17**, 828–839 (2007).
69. Aide, T. M., Zimmerman, J. K., Pascarella, J. B., Rivera, L. & Marciano-Vega, H. Forest regeneration in a chronosequence of tropical abandoned pastures: implications for restoration ecology. *Restor. Ecol.* **8**, 328–338 (2000).



**Extended Data Figure 1 | Relative recovery of AGB after 20 years in relation to abiotic factors, forest cover, and land use.** a, Annual precipitation; b, CWD; c, rainfall seasonality; d, CEC; e, percentage forest cover in the surrounding matrix; f, previous land use (SC, shifting cultivation,  $N = 17$ ; SC & PA, some plots shifting cultivation, some plots

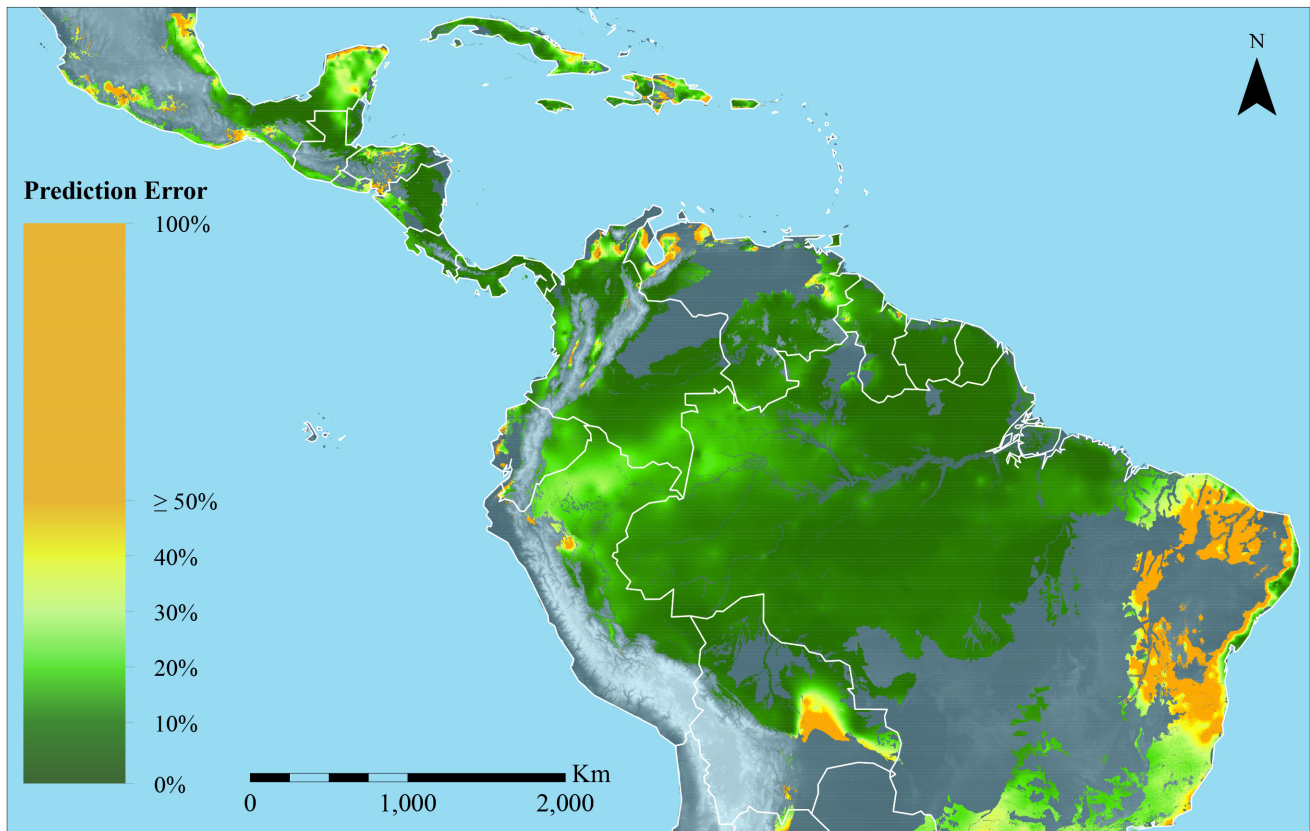
pasture,  $N = 2$ ; PA, pasture,  $N = 9$ ; means  $\pm$  s.e.m. are shown). Relative recovery is expressed as the ratio of AGB after 20 years over median AGB of old-growth forest (as a percentage). Regression lines are shown while keeping the other variable constant at the mean value across sites ( $P = 0.040$  for  $1/\text{rainfall}$ ,  $P = 0.027$  for CEC,  $R^2 = 0.23$ ,  $N = 28$  Neotropical forest sites).



**Extended Data Figure 2 | AGB recovery after 20 years in relation to abiotic factors, forest cover, and land use.** a, Rainfall seasonality; b, CEC; c, percentage forest cover in the surrounding matrix; d, previous land use (SC,  $N = 19$ ; SC & PA,  $N = 9$ ; PA,  $N = 15$ ; means  $\pm$  s.e.m. are shown).

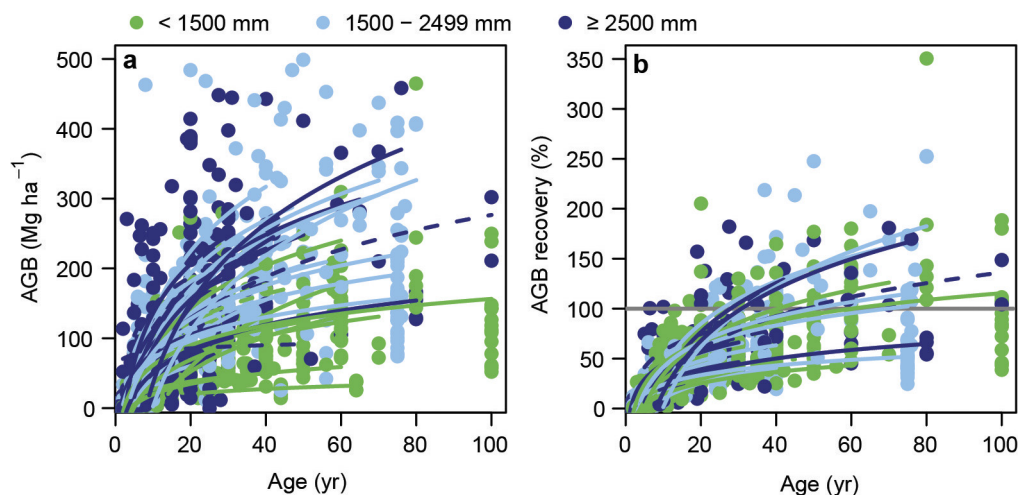
For rainfall seasonality, the regression line is shown based upon the multiple regression model that also includes rainfall and CWD, and where these variables were kept constant at the mean value across sites (two-sided  $P = 0.003$ , see Fig. 2 for these models for rainfall and CWD).





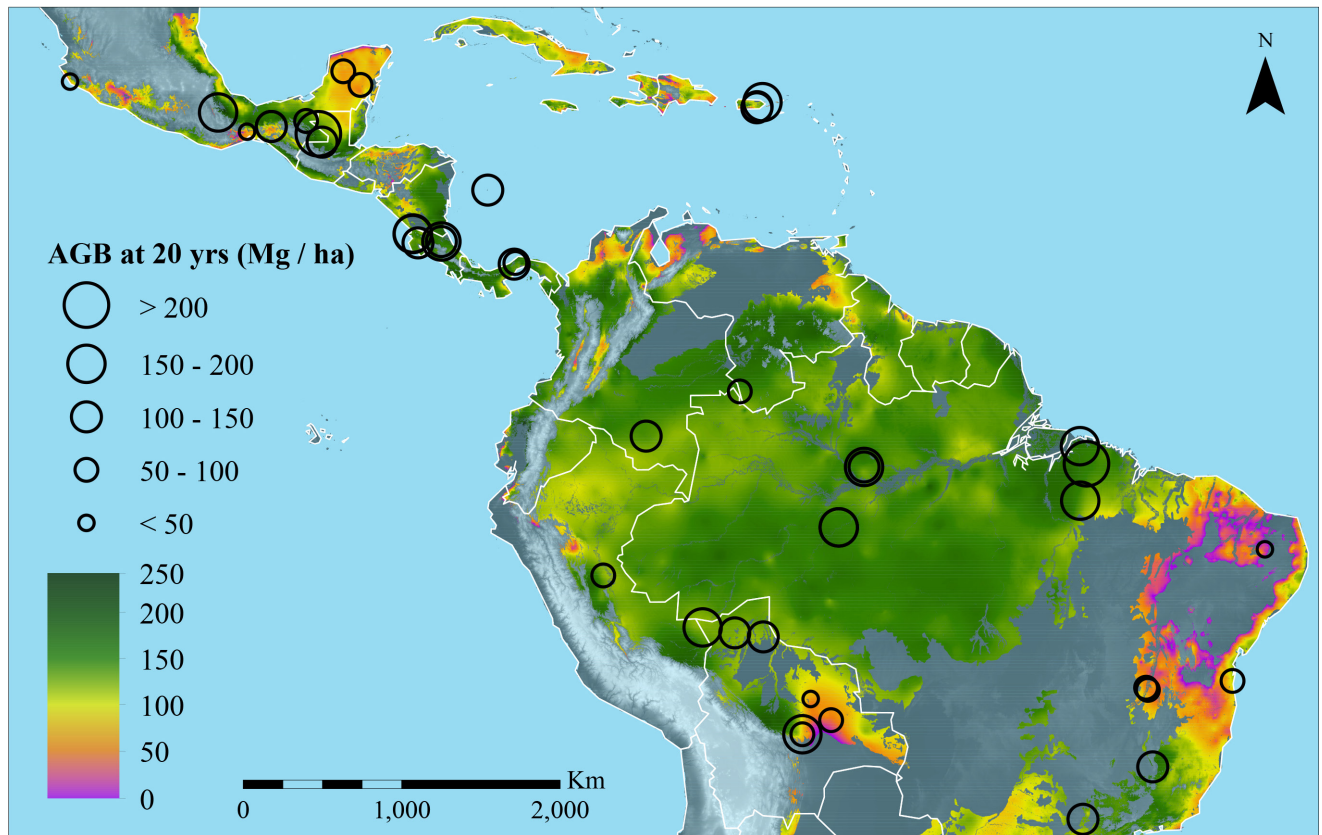
**Extended Data Figure 3 | Uncertainty map of potential biomass recovery of Neotropical secondary forests.** The uncertainty is based on the 95% confidence interval of the mean predicted AGB after 20 years (see Fig. 3 and Methods). It is expressed as a percentage of the predicted AGB:  $100 \times (0.5 \times 95\% \text{ confidence interval of the mean}) / \text{predicted AGB}$ .

In general the uncertainty is low: 80.32% of the mapped area has an uncertainty less than 20%, and 10.2% of the mapped area has an uncertainty between 20% and 30%. Because it is a relative uncertainty, it is highest in the driest areas, which have a low predicted biomass.



**Extended Data Figure 4 | Relationship between forest biomass and stand age using chronosequence studies in Neotropical secondary forest sites. a, AGB ( $N = 44$ ); b, AGB recovery ( $N = 28$ ).** The same as Fig. 1 but with plots and regression lines coloured by forest type: green, dry forest (<1,500 mm rainfall per year); light blue, moist forest (1,500–2,499 mm yr<sup>-1</sup>); dark blue, wet forest (≥2,500 mm yr<sup>-1</sup>). Each line represents a different chronosequence. The original plots on which the

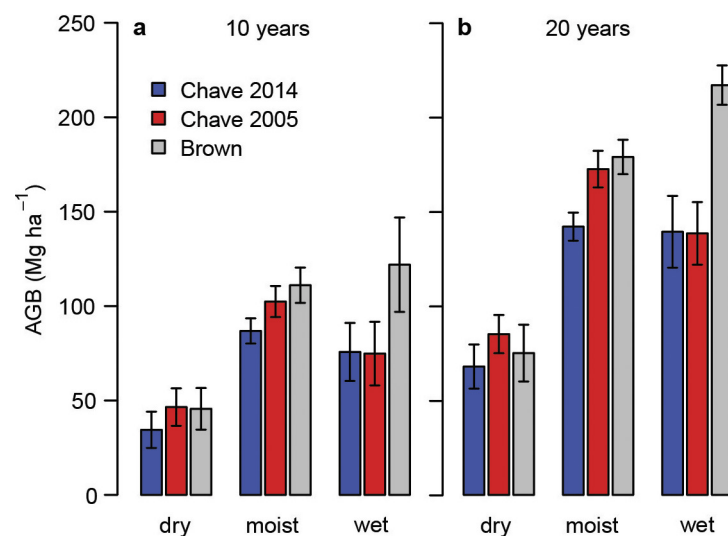
regression lines are based are shown ( $N = 1,364$  for AGB,  $N = 995$  for AGB recovery). AGB recovery is defined as the AGB of the secondary forest plot compared with the median AGB of old-growth forest plots in the area, multiplied by 100. Significant relations (two-sided  $P \leq 0.05$ ) are indicated by continuous lines, non-significant relationships (two-sided  $P > 0.05$ ) are indicated by broken lines. Plots of 100 years old are also second-growth.



**Extended Data Figure 5 | Potential biomass recovery map of Neotropical secondary forests.** The same as Fig. 3 but with colour-blind-friendly colour coding. The total potential AGB accumulation over 20 years of lowland secondary forest growth was calculated on the basis of a regression equation relating AGB with annual rainfall ( $AGB = 135.17 - 103,950 \times 1/rainfall + 1.522 \times rainfall$

seasonality +  $0.1148 \times CWD$ ; see Methods). The colour indicates the amount of forest cover recovery (purple, low recovery; green, high recovery). The 44 study sites are indicated by circles; the size of the symbols scales with the AGB attained after 20 years. The grey areas do not belong to the tropical forest biome. The map focuses on lowland tropical forest (altitude < 1,000 m).





**Extended Data Figure 6 | AGB of secondary forest.** a, AGB 10 years and b, 20 years after land abandonment. Predicted mean AGB is given for three different forest types (dry (<1,500 mm rainfall), moist (1,500–2,499 mm),

wet ( $\geq 2,500$  mm)) using three different allometric equations (indicated by different colours). These allometric equations are ordered from left to right as ref. 34 (blue), ref. 33 (red), and ref. 32 (grey). Means  $\pm$  s.e.m. are shown.

Extended Data Table 1 | Overview of the sites included in the study

Site	Country	Lat.	Long.	Rainfall	CWD	RS	CEC	FC	LU	AGB	%AGB	Age	Ref.
Bolpebra	Bolivia	-11.44	-69.16	1800	-301	55	6.4	80	SC	193.2	63.5	4-47	44
El Tigre	Bolivia	-11.98	-65.72	1780	-375	66	5.8	90	SC	145.0	32.3	3-25	45
El Turi	Bolivia	-11.75	-67.33	1833	-288	58	5.8	95	SC	129.4	47.8	2-40	45
Salvatierra	Bolivia	-15.50	-63.03	1200	-642	61	4.0	70	SC	38.3	29.5	1-36	46
San Lorenzo	Bolivia	-16.70	-61.87	1129	-718	59	4.5	40	SC	82.8	64.5	3-50	47
Surutu 1	Bolivia	-17.50	-63.50	1600	-496	54	6.1	45	SC	60.5	-	4-20	48
Surutu 2	Bolivia	-17.50	-63.50	1600	-496	54	6.1	45	SC	154.3	-	6-25	48
Bahia	Brazil	-14.48	-39.09	2000	-7	17	6.6	80	SC	96.7	25.4	10-40	49
Cajueiro	Brazil	-14.98	-43.90	840	-907	96	13.8	80	PA	52.2	-	15-59	
Serra do Cipo	Brazil	-19.35	-43.62	1519	-341	86	5.4	60	PA	130.3	-	8-50	
Eastern Pará 1	Brazil	-4.26	-47.73	1898	-548	88	7.8	50	SC & PA	168.2	-	5-25	50
Eastern Pará 2	Brazil	-2.16	-47.38	2460	-274	64	4.6	52	SC & PA	224.7	-	5-40	50
Eastern Pará 3	Brazil	-1.17	-47.75	2785	-306	68	4.3	58	PA	176.5	-	2-70	50
Manaus ( <i>Cecropia</i> pathway)	Brazil	-2.35	-60.00	2400	-16	32	2.0	100	SC	181.2	55.8	5-31	51
Manaus ( <i>Vismia</i> pathway)	Brazil	-2.35	-60.00	2400	-16	32	2.0	100	PA	140.6	43.3	2-25	51
Mata Seca	Brazil	-14.86	-43.99	825	-914	95	3.5	80	PA	58.8	-	14-58	52
Patos	Brazil	-7.02	-37.25	750	-1276	116	11.2	60	SC & PA	20.3	-	22-64	53
Rio Madeira (anthropogenic soil)	Brazil	-5.77	-61.43	2507	-109	46	3.4	79	SC	187.9	74.4	5-30	54
Rio Madeira (oxisol)	Brazil	-5.77	-61.43	2507	-109	46	4.7	79	SC	159.2	56.2	6.5-30	54
Sao Paulo	Brazil	-22.32	-47.57	1367	-297	72	10.2	31	PA	123.2	35.7	11-45	
Aracuaia	Colombia	-0.60	-72.37	3059	-13	29	9.7	95	SC	134.5	39.2	2-30	55
Providencia Island	Colombia	13.35	-81.35	1584			31.8	60	SC & PA	135.9	-	6-56	56
San Carlos	Colombia	1.93	-67.05	3565	0	23	5.8	80	SC	84.6	35.5	9-80	57
Santa Rosa (oak forest)	Costa Rica	10.89	-85.60	1765	-608	91	32.4	70	PA	161.3	-	5-70	58
Palo Verde	Costa Rica	10.36	-85.31	1444	-622	88	39.0	70	PA	142.9	-	7-60	58
Santa Rosa	Costa Rica	10.86	-85.61	1765	-621	93	32.4	70	PA	170.9	84.9	6-70	58
Sarapiquí (Chazdon)	Costa Rica	10.43	-83.98	4000	0	29	29.0	77	PA	171.5	59.6	10-41	59
Sarapiquí (Letcher)	Costa Rica	10.43	-83.98	4000	0	29	29.0	77	PA	145.4	76.6	10-42	60
Chajul	Mexico	16.07	-90.75	3000	-248	62	10.8	50	SC	105.3	59.4	0-27	61
Chamela	Mexico	19.50	-105.05	788	-905	109	21.0	60	PA	34.8	49.9	3-15	62
Chinantla	Mexico	17.75	-96.65	3700	-596	83	10.1	95	SC	155.7	63.8	5-50	
El Ocote 1	Mexico	16.96	-93.63	1500	-558	69	26.3	80	SC	101.4	33.6	2-75	63
El Ocote 2	Mexico	16.96	-93.63	2000	-558	69	18.6	80	SC	116.4	62.3	3-75	63
Lacandona	Mexico	16.54	-90.96	2500	-325	68	37.0	87	SC & PA	206.3	-	1-31	
Marqués de Comillas	Mexico	16.33	-90.67	2250			21.4	49	SC & PA	206.3	-	2-50	
Nizanda	Mexico	16.66	-95.01	878	-1036	102	30.6	64	SC	34.7	25.4	2-60	64
Quintana Roo	Mexico	19.31	-88.57	1250	-508	54	64.6	85	SC & PA	90.0	66.4	2-100	
Yucatán	Mexico	20.08	-89.55	1100	-601	65	39.9	94	SC	69.3	66.7	3-70	65
Zona Norte	Mexico	17.27	-91.66	2750	-213	56	37.0	67	SC & PA	85.6	-	2-52	
Agua Salud	Panama	9.22	-79.78	2700	-326	57	9.3	70	SC & PA	81.8	-	2-31	66
Barro Colorado Island	Panama	9.15	-79.85	2600	-319	57	9.3	100	SC & PA	120.5	59.3	20-100	67
Pucallpa	Peru	-8.50	-74.80	1570	-56	34	21.4	43	SC & PA	99.2	-	5-30	
Cayey	Puerto Rico	18.02	-66.08	2000	-220	41	23.9	50	PA	110.5	61.8	10-80	68
El Carite	Puerto Rico	18.08	-66.07	1822	-165	35	20.5	71	PA	127.7	66.7	4-77	69
Luquillo	Puerto Rico	18.34	-65.76	2660	-12	31	9.5	75	PA	150.1	67.8	9-76	69

An overview of the name, country, latitude (Lat.) and longitude (Long.), rainfall (in millimetres), rainfall seasonality (RS), CWD (in mm per year), CEC (in centimoles of positive charge per kilogram of soil), percentage forest cover in the landscape matrix (FC; as a percentage), prior land use (LU: SC, shifting cultivation; SC & PA, some plots shifting cultivation, some plots pasture; PA, pasture), AGB (in megagrams per hectare) after 20 years, relative biomass recovery after 20 years (%AGB, as a percentage), the minimum and maximum secondary forest age (in years) included in the chronosequence, and a reference (refs 44–69) for each site is given.

**Extended Data Table 2 | Overview of the modelling results of absolute ( $N=43$ , one site was excluded because of missing climatic data) and relative ( $N=28$ ) AGB recovery after 20 years in relation to rainfall, CEC, land use, and forest cover in the landscape matrix**

Variable	Model	$\Delta\text{AIC}_c$	$R^2$
AGB at 20 yr (Mg/ha)	1/rainfall + CWD + Rainfall seasonality	0.00	0.59
AGB at 20 yr (%)	1/rainfall + CEC	0.00	0.23
	1/rainfall + CEC + Rainfall seasonality	1.53	0.27

All models within two  $\text{AIC}_c$  units of the best model are given. The change in AIC for small sample sizes compared with the best model ( $\Delta\text{AIC}_c$ ) and  $R^2$  are given.



# New geological and palaeontological age constraint for the gorilla–human lineage split

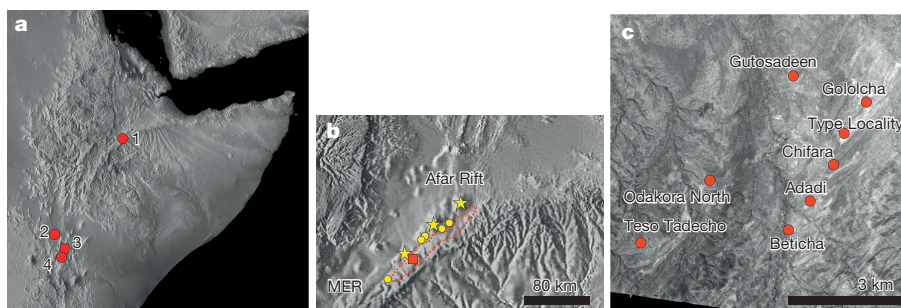
Shigehiro Katoh<sup>1</sup>, Yonas Beyene<sup>2,3</sup>, Tetsumaru Itaya<sup>4</sup>, Hironobu Hyodo<sup>4</sup>, Masayuki Hyodo<sup>5</sup>, Koshi Yagi<sup>6</sup>, Chitaro Gouzu<sup>6</sup>, Giday WoldeGabriel<sup>7</sup>, William K. Hart<sup>8</sup>, Stanley H. Ambrose<sup>9</sup>, Hideo Nakaya<sup>10</sup>, Raymond L. Bernor<sup>11</sup>, Jean–Renaud Boisserie<sup>3,12</sup>, Faysal Bibi<sup>13</sup>, Haruo Saegusa<sup>14</sup>, Tomohiko Sasaki<sup>15</sup>, Katsuhiko Sano<sup>15</sup>, Berhane Asfaw<sup>16</sup> & Gen Suwa<sup>15</sup>

The palaeobiological record of 12 million to 7 million years ago (Ma) is crucial to the elucidation of African ape and human origins, but few fossil assemblages of this period have been reported from sub-Saharan Africa. Since the 1970s, the Chorora Formation, Ethiopia, has been widely considered to contain ~10.5 million year (Myr) old mammalian fossils<sup>1–7</sup>. More recently, *Chororapithecus abyssinicus*, a probable primitive member of the gorilla clade<sup>6</sup>, was discovered from the formation. Here we report new field observations and geochemical, magnetostratigraphic and radioisotopic results that securely place the Chorora Formation sediments to between ~9 and ~7 Ma. The *C. abyssinicus* fossils are ~8.0 Myr old, forming a revised age constraint of the human–gorilla split. Other Chorora fossils range in age from ~8.5 to 7 Ma and comprise the first sub-Saharan mammalian assemblage that spans this period. These fossils suggest indigenous African evolution of multiple mammalian lineages/groups between 10 and 7 Ma, including a possible ancestral-descendent relationship between the ~9.8 Myr old *Nakalipithecus nakayamai*<sup>8</sup> and *C. abyssinicus*. The new chronology and fossils suggest that faunal provinciality between eastern Africa and Eurasia had intensified by ~9 Ma, with decreased faunal interchange thereafter<sup>9–12</sup>. The Chorora evidence supports the hypothesis of *in situ* African evolution of the *Gorilla*–*Pan*–human clade, and is concordant with the deeper divergence estimates of humans and great apes based on lower mutation rates of  $\sim 0.5 \times 10^{-9}$  per site per year (refs 13–15).

The Chorora Formation type locality fauna<sup>1–3</sup> has been known as one of few key windows revealing Late Miocene mammalian evolution in sub-Saharan Africa. In 2006/07 the fossil ape *C. abyssinicus*<sup>6</sup> was discovered at Beticha, a locality ~3 km from the type locality (Fig. 1). Because *Chororapithecus* is considered a strong candidate for membership in the gorilla clade<sup>6</sup>, its age is important in temporally constraining the gorilla–human species divergence<sup>13,14</sup>. The probable cladistic position of *Chororapithecus* (Extended Data Fig. 1) is based on a suite of derived molar features shared with gorillas, thought to indicate an early phase of increased herbivory<sup>6</sup>.

The >10 Myr age of the Chorora Formation was proposed on the basis of stratigraphic and geochronological frameworks established in the 1970s<sup>1,2</sup>. In the 1990s and 2000s, additional radioisotopic dating and biochronological considerations refined this chronology<sup>3,4</sup>. In 2007, the *Chororapithecus*-bearing strata were correlated to the Chorora type locality sequence and thereby thought to be older than 10 Ma (ref. 6). Here we present chronostratigraphic data from new field research, and show that the Chorora Formation sediments are predominantly younger than 9 Ma. We also present an overview of the newly recovered Chorora Formation fossils, which fill the gap between sub-Saharan fossils that are either older than 9–9.5 Ma (refs 7, 8, 16) or younger than 7–7.4 Ma (refs 17, 18). The Chorora fossils thus provide new insights into mammalian evolution in Africa, in a period crucial to elucidating the emergence of the African ape and human clade<sup>5</sup>.

The Chorora Formation extends northeast to southwest for >100 km at the transition zone between the southern Afar Rift and the northern



**Figure 1 | Location of Chorora Formation sedimentary exposures.** **a**, Digital elevation map of eastern Africa and Late Miocene faunal sites. 1, Chorora; 2, Lothagam; 3, Samburu Hills; 4, Nakali. **b**, Transition zone of the Main Ethiopian Rift (MER) and the Afar Rift. Dotted line indicates extent of Chorora Formation sediments<sup>1,2</sup>. Yellow stars and circles are

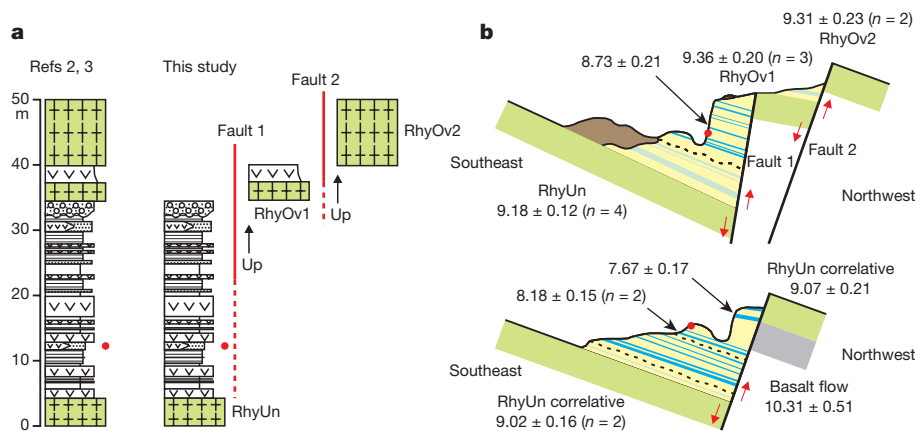
5–7 Myr old volcanic centres (northeast to southwest: Afdem, Asabot and Gumbi) and flows<sup>19,21</sup>. Red square shows location of **c**. **c**, Chorora Formation exposures and localities. **a**, **b**, Images from the National Geophysical Data Center; **c**, air photograph provided by the Ethiopian Mapping Agency.

<sup>1</sup>Division of Natural History, Hyogo Museum of Nature and Human Activities, Sanda 669-1546, Japan. <sup>2</sup>Association for Conservation of Culture Awassa, PO Box 6686, Addis Ababa, Ethiopia.

<sup>3</sup>Centre français des études éthiopiennes (CFEE), USR CNRS 3137, French Ministry for Foreign Affairs, PO Box 5554, Addis Ababa, Ethiopia. <sup>4</sup>Research Institute of Natural Sciences, Okayama University of Science, Okayama 700-0005, Japan. <sup>5</sup>Research Center for Inland Seas, Kobe University, Kobe 657-8501, Japan. <sup>6</sup>Hiruzen Institute for Geology and Chronology, Okayama 703-8252, Japan. <sup>7</sup>EES-14/MS D462, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. <sup>8</sup>Department of Geology and Environmental Earth Science, Miami University, 133 Culler Hall, Oxford, Ohio 45056, USA. <sup>9</sup>Department of Anthropology, University of Illinois, Urbana, Illinois 61801, USA. <sup>10</sup>Department of Earth and Environmental Sciences, Kagoshima University, Kagoshima 890-0065, Japan. <sup>11</sup>Department of Anatomy, Howard University, Washington DC 20059, USA. <sup>12</sup>Institut de Paléoprimatologie, Paléontologie Humaine : Évolution et Paléoenvironnements (IPHEP), UMR CNRS 7262, Université de Poitiers, 86022 Poitiers, France. <sup>13</sup>Museum für Naturkunde - Leibniz Institute for Evolution and Biodiversity Science, Invalidenstraße 43, 10115 Berlin, Germany.

<sup>14</sup>Institute of Natural and Environmental Sciences, University of Hyogo, Sanda 669-1546, Japan. <sup>15</sup>The University Museum, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.

<sup>16</sup>Rift Valley Research Service, PO Box 5717, Addis Ababa, Ethiopia.



**Figure 2 | Summary of Chorora Formation fault block relationships.** **a**, Schematic comparison of hypotheses. Left: rhyolitic ignimbrite units above and below the Chorora Formation sediments. Right: successive fault blocks with single rhyolitic ignimbrite unit. **b**, Fault block relationships

Main Ethiopian Rift<sup>1,2,19</sup> (Fig. 1). The formation overlies Miocene basaltic lavas (Trap Series) ranging from >23 to 9 Ma, with intercalated silicic units<sup>1,19–21</sup>. Recent studies suggest that a topographically low pre-rift landscape prevailed in this region until ~11 to 10 Ma (refs 19, 20). The Chorora Formation accumulated in a marginal graben formed along the emerging southeastern Afar Rift border fault as extension between the Nubian and Somali plates accelerated in the Late Miocene<sup>20</sup>. We show that lacustrine and fluvial sedimentation of the Chorora Formation occurred between 9.2 and ~7 Ma. Thereafter, substantial rift floor subsidence occurred further northwest, as inferred from the chain of volcanic centres and widely distributed volcanic flows dated between ~7 and 5 Ma (refs 19–21).

The ~10.5 Myr age formerly attributed to the Chorora Formation type locality fossils was based on stratigraphy and geochronology conducted by previous researchers<sup>1–4</sup>. At the type locality and adjacent areas, sedimentary packages were assigned to ‘series 1’ and ‘series 2’ in ascending stratigraphic order<sup>2,3</sup>. The type locality fossils were collected from a sandy unit of the ‘series 2’ sediments that was assumed to underlie rhyolitic ignimbrite units dated to ~10 Ma (refs 1–3).

We focused on refining the chronostratigraphy of both the type locality and the *Chororapithecus*-bearing Beticha locality. Because the Beticha and type locality sections differ in the represented sedimentary units<sup>3,6</sup>, we investigated a range of localities in establishing a more comprehensive Chorora Formation chronostratigraphic framework. Despite variation among localities, we observed a distinct fine-grained welded rhyolitic ignimbrite unit to consistently underlie the Chorora Formation sediments. These localities occur as northeast–southwest trending, variably (mostly antithetically) tilted fault blocks, often resulting in the underlying resistant volcanic strata elevated relative to the adjacent fault block (Fig. 2). At the type locality, we could not identify a section where the rhyolitic ignimbrite unambiguously overlies the Chorora Formation sediments. Instead, we observed sections where the rhyolitic ignimbrite unit previously considered to overlie the Chorora Formation sediments (hereafter RhyOv1) occurs in fault contact with the uppermost ‘series 2’ sediments (Extended Data Fig. 2). A possibly second ignimbrite (RhyOv2) occurs ~20 m topographically higher than RhyOv1 and forms the local ridge top >200 m west of the main type locality exposures. We hypothesized that both RhyOv1 and RhyOv2 are step-faulted occurrences of the rhyolitic ignimbrite unit (RhyUn) that directly underlies the Chorora Formation sediments. Previously, these flows yielded whole-rock K–Ar ages<sup>3</sup> of  $9.84 \pm 0.6$  (RhyOv2),  $10.1 \pm 0.6$  (RhyOv1) and  $11.0 \pm 0.9$  (RhyUn) Ma (here and elsewhere errors are  $2\sigma$ ). To verify these dates using current K–Ar techniques (see Methods), we analysed chemically homogeneous 100–200 mesh (150–74  $\mu$ m) feldspar separates. The K–Ar ages, feldspar chemistry and remanent

at the type locality (upper) and Beticha (lower). Numbers are K–Ar ages and errors ( $2\sigma$ ) in millions of years. Weighted means are shown with number of samples in parenthesis. Red dots, fossil horizons; dotted lines, depositional hiatus.

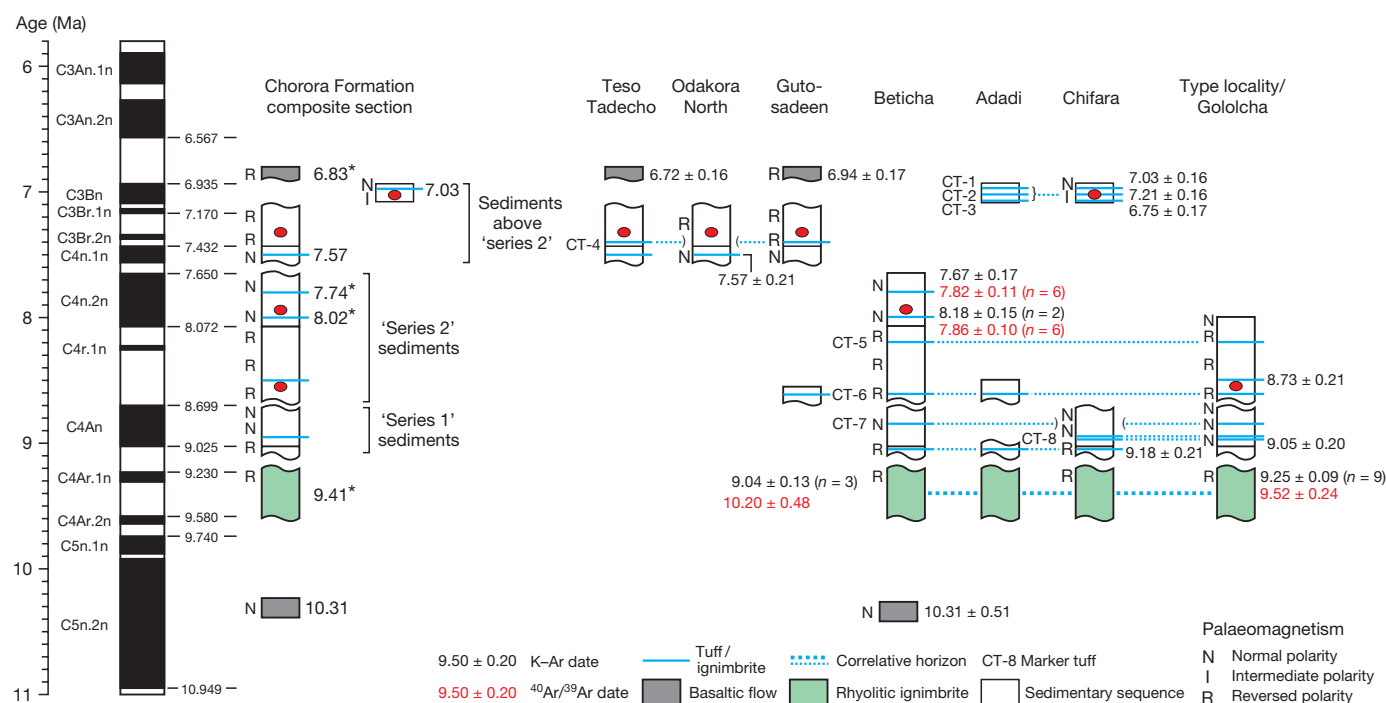
magnetism were compared across samples to test the hypothesis that RhyOv1, RhyOv2 and RhyUn are correlative.

We sampled the type locality RhyUn ( $n = 4$ , weighted mean age of  $9.18 \pm 0.12$  Ma) at the same section as previous workers, and the RhyOv1 and RhyOv2 at exposures continuous with the previous sample locations (Extended Data Fig. 2). These yielded K–Ar ages of  $9.36 \pm 0.20$  (RhyOv1,  $n = 3$ ) and  $9.31 \pm 0.23$  (RhyOv2,  $n = 2$ ) Ma, respectively, slightly older than—but statistically indistinguishable from—the RhyUn dates (Extended Data Table 1). Feldspar chemistry is comparable across sections and samples (Extended Data Fig. 3), suggesting a single flow unit or closely spaced successive units. All measured RhyUn, RhyOv1 and RhyOv2 samples ( $n = 5$ ) exhibited reverse magnetization. Our results indicate that the hypothesis of RhyUn, RhyOv1 and RhyOv2 equivalency cannot be rejected. We also investigated rhyolitic ignimbrite sections at Beticha that are lithologically and stratigraphically indistinguishable from the type locality sections (Fig. 2). We analysed three rhyolitic ignimbrite samples from Beticha. These resulted in K–Ar ages (weighted mean age:  $9.04 \pm 0.13$  Ma), feldspar chemistry and palaeomagnetism concordant with the type locality RhyUn/Ov1/Ov2.

The weighted mean age of 12 RhyUn/Ov1/Ov2 and correlative samples is  $9.17 \pm 0.08$  Ma. However, the individual K–Ar ages varied from  $8.76 \pm 0.21$  to  $10.08 \pm 0.45$  Ma, suggesting uncertainties exceeding experimental error. Step-heating  $^{40}\text{Ar}$ – $^{39}\text{Ar}$  analysis of feldspars resulted in an older weighted mean age of  $9.66 \pm 0.22$  Ma ( $n = 2$ ) (Extended Data Fig. 4 and Extended Data Table 1). Combined with the consistently reverse magnetization of RhyUn/Ov1/Ov2, these dates constrain the unit within Chron C4Ar (9.03–9.74 Ma) (Fig. 3).

The above results and other analyses enable construction of a composite Chorora Formation stratigraphy across eight fossiliferous localities (Fig. 3 and Extended Data Figs 5 and 6). At the type locality, the >20 m thick ‘series 2’ section that stratigraphically overlies the RhyUn consistently exhibits reverse polarity. A tuff from the middle of this interval, just above the main vertebrate fossil horizon, yielded a K–Ar age of  $8.73 \pm 0.21$  Ma. This suggests that the long type locality interval of reverse magnetism corresponds to Chron C4r–2r (8.25–8.70 Ma) and that the type locality fossils are ~8.5 Myr old.

At Beticha, a gravelly sand unit containing the *Chororapithecus* fossils occurs in the upper portion of a >40 m thick sequence, just above a pumiceous tuff dated to 8.02 Ma (average of  $8.18 \pm 0.15$  by K–Ar and  $7.86 \pm 0.10$  by  $^{40}\text{Ar}$ – $^{39}\text{Ar}$ ). An ~5 m predominantly silty interval overlies this fossil-bearing unit. This is capped by a consolidated tuff dated to 7.74 Ma (average of  $7.67 \pm 0.17$  by K–Ar and  $7.82 \pm 0.11$  by  $^{40}\text{Ar}$ – $^{39}\text{Ar}$ ). A magnetic transition occurs just below the 8.02 Myr old tuff. The overlying sediments are normally magnetized, suggesting that this transition corresponds to the C4r–C4n boundary (8.072 Ma).



**Figure 3 | Chronostratigraphy of the Chorora Formation.** Schematic summary across eight localities. The type locality and Gololcha sections are shown in the same column. Red ovals, fossil horizons. Numbers are radioisotopic dates and errors ( $2\sigma$ ) in millions of years; weighted means

are shown with number of samples in parenthesis. The composite section shows summary ages: asterisks indicate average K–Ar ages, or averages of K–Ar and  $^{40}\text{Ar}$ – $^{39}\text{Ar}$  ages (using weighted means).

Interpolating between this and the capping 7.74 Ma age yields an 8.0 Ma estimate for the *Chororapithecus* and associated fossils.

At the type locality, Beticha and elsewhere, the upper 'series 2' or higher sediments are of variable lithology and age. At Chifara, a deep 'series 1' sequence is in fault contact with an upper tuffaceous/fluvial sequence containing a fossil-bearing unit. The K–Ar ages and normal magnetization suggest a 7.0–7.15 Ma age of the fossils. At three other newly established localities (Gutosadeen, Odakora North and Teso Tadecho), mammalian fossils occur in sands intercalated in floodplain silts that overlie an ignimbrite with a K–Ar age of  $7.57 \pm 0.21$  Ma and underlie basalts with ages of  $6.72 \pm 0.16$  and  $6.94 \pm 0.17$  Ma (Fig. 3).

Finally, we tested the above-established Chorora Formation chronostratigraphy by comparing the Chorora fossils with those of preceding and succeeding eastern African assemblages, in particular those of Nakali/Samburu Hills (9.9–9.5 Ma) and Lothagam Lower Nawata Member (7.4–6.5 Ma) of Kenya (Extended Data Table 2). The descriptive details of the Chorora Formation fauna are presented elsewhere<sup>22</sup>. The Chorora type locality and Beticha assemblages share archaic taxa with Nakali/Samburu Hills, but also have distinctly derived elements. Notable first appearances include a low-crowned primitive species of hippopotamus at Beticha. This new taxon morphologically bridges *Kenyapotamus* and the more advanced Hippopotaminae that comprise the post-7.4 Ma hippopotamid diversity<sup>10</sup>. A primitive-looking large bovid premolar may represent an early bovin or boselaphin. The ~8.5–8 Ma Chorora cercopithecoid teeth include at least two species of colobine and a cercopithecine. The latter appears to be a small papionin, so far the earliest recognized record of Cercopithecinae. The dominant colobine is larger than the 10–9 Myr old Kenyan *Microcolobus*<sup>8,23</sup>, and has conservative molar cusp notches as in Late Miocene Eurasian *Mesopithecus* spp. Another taxon is similar in molar size to *Microcolobus tugenensis* but differs in details. A third colobine may be represented by a mandibular molar that is higher crowned with a deep lingual notch well within modern colobine ranges of variation<sup>17,24</sup>. These new cercopithecoid fossils suggest considerable hidden diversity in the fossil-poor early to middle Late Miocene of Africa. The Chorora hipparionins suggest that two lineages, a larger and smaller

species, existed and went through a mostly *in situ* African transformation during the Late Miocene<sup>22</sup>: from the ~9.5 Ma Samburu Hills '*Cormohipparion*' aff. *africanum* (most likely two species), through the Chorora hipparionins, to the post-7.4 Ma *Eurygnathohippus turkanensis* and *E. feibeli*. The newly recognized upper Chorora fauna, estimated at 7.5–7.0 Ma, shares more taxa with the Lothagam Lower Nawata Member.

The Chorora fossils suggest that multiple indigenous African mammalian lineages developed between ~9.5 and 7 Ma, including hippopotamines, nyanzachoere suids, bovids, hipparionin equids and cercopithecids. However, inter-relationships with Eurasia must also be considered. In particular, the hipparionin evidence shows a parallel appearance of large robust forms across continents: the Samburu Hills/Chorora/Lothagam 9.6–7 Ma sub-Saharan African sequence and the <9–7 Ma Siwalik *Sivalhippus* spp. It is probable that eastern Africa had a closer bioprovincial relationship with southern Asia than with western Eurasia<sup>11,25</sup>, and that population and species lineage interchange intermittently occurred between continents. In sub-Saharan Africa, the immigrant hipparionins became the first to extensively exploit  $C_4$  grasses as early as ~9.6 Ma in Kenya<sup>26</sup> and by ~8.5 Ma at Chorora<sup>22</sup>. The emerging diversity of African cercopithecoid monkeys in the Late Miocene implies that from such a pool of diversity only restricted lineages actually extended into Eurasia, for example, the semi-terrestrial colobine *Mesopithecus* (~8.5–7 Ma) of the Eurasian open-woodland 'Pikermian paleobiome' (~9.5–6.5 Ma)<sup>27</sup>. Finally, the early *Alilepus* recorded at Chorora (most probably 7.5–7 Ma) indicates that the 'leporid event'<sup>28</sup> penetration into Africa might have occurred near-simultaneously with southern Asia (recorded at 7.4 Ma). These Afro-Eurasian links suggest an increasingly filtered faunal exchange after ~9 Ma (refs 9–12).

The *Chororapithecus*-bearing Beticha fauna is distinct from that of the type locality in taxonomic relative abundances<sup>6,22</sup>, exhibiting clear dominance of primates, hippopotamids and browsing ungulates. Enamel isotopic analyses indicate that these early hippopotamids (most abundant next to primates at Beticha) appear to have been near-pure  $C_3$  feeders ( $\delta^{13}\text{C} \leq -12.0$ ; Extended Data Table 3). Combined with the



very low  $\delta^{18}\text{O}$  values of the two individuals, the data suggest that the primitive Beticha hippopotamids fed in humid, closed habitats that lacked  $\text{C}_4$  plants. Although more evidence is needed, it is probable that, especially after  $\sim 9$  Ma, faunal interchange between Eurasia and sub-Saharan Africa was increasingly limited, with potential routes not conducive to dispersal of closed-habitat-dependent species lineages, including those of primates.

Our new age-estimate of *Chororapithecus* at  $\sim 8$  Ma revises its chronocladistic relationship with *Nakalipithecus* (Extended Data Fig. 1). It provides a minimum age of the gorilla–human species lineage split at either  $\sim 8$  or  $\sim 10$  Ma, depending on the cladistic position of *Nakalipithecus*. The Kenyan *N. nakayamai* lacks some of the derived molar features of *C. abyssinicus*, but shares with *C. abyssinicus* large postcanine size and possibly a prominent transverse crest in the lower molars (inferred from wear pattern). Thus, *N. nakayamai* may represent a stem member of the gorilla clade more primitive than *C. abyssinicus*, or precede the gorilla split as previously suggested<sup>8</sup>.

Until recently, many researchers preferred a relatively recent timing of modern great ape and human divergences, a 4–6 Ma split for humans and chimpanzees and a  $\leq 8$  Ma split for humans and gorillas<sup>5,13,29</sup>. However, genomic modelling of modern great ape and human divergence timings based on a comparatively low intergenerational mutation rate<sup>13–15</sup> are in accord with the fossil evidence. The Chorora evidence effectively refutes a 9–10 Ma Eurasian origin of the African ape and human clade<sup>30</sup>, and highlights the need of more fossils, especially those of large-bodied apes from the 12 to 7 Ma time interval of sub-Saharan Africa<sup>5,8,30</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 27 September; accepted 3 December 2015.**

- Sickenberg, O. & Schönfeld, M. in *Afar Depression of Ethiopia* (eds Pilger, A., Rössler, A.) 277–284 (Schweizerbart, 1975).
- Tiercelin, J. J., Michaux, J. & Bandet, Y. Le Miocène supérieur du sud de la dépression de l'Afar, Éthiopie; sédiments, faunes, âges isotopiques. *Bull. Soc. Geol. Fr. Ser. 7* **XXI**, 255–258 (1979).
- Geraads, D., Alemseged, Z. & Bellon, H. The late Miocene mammalian fauna of Chorora, Awash basin, Ethiopia: systematics, biochronology and the  $^{40}\text{K}$ – $^{40}\text{Ar}$  ages of the associated volcanics. *Tertiary Res.* **21**, 113–127 (2002).
- Bernor, R. L., Kaiser, T. M. & Nelson, S. V. The oldest Ethiopian hipparion (Equinae, Perissodactyla) from Chorora: systematics, paleodiet and paleoclimate. *Cour. Forsch. Senckenberg* **246**, 213–226 (2004).
- Cote, S. M. Origins of the African hominoids: an assessment of the palaeobiogeographical evidence. *C. R. Palevol* **3**, 323–340 (2004).
- Suwa, G., Kono, R. T., Katoh, S., Asfaw, B. & Beyene, Y. A new species of great ape from the late Miocene epoch in Ethiopia. *Nature* **448**, 921–924 (2007).
- Werdelin, L. & Sanders, W. J. *Cenozoic Mammals of Africa* (Univ. California Press, 2010).
- Kunimatsu, Y. *et al.* A new Late Miocene great ape from Kenya and its implications for the origins of African great apes and humans. *Proc. Natl Acad. Sci. USA* **104**, 19220–19225 (2007).
- Douady, C. J., Catzeflis, F., Raman, J., Springer, M. S. & Stanhope, M. J. The Sahara as a vicariant agent, and the role of Miocene climatic events, in the diversification of the mammalian order Macroscelidea (elephant shrews). *Proc. Natl Acad. Sci. USA* **100**, 8325–8330 (2003).
- Boisserie, J.-R., Fisher, R. E., Lihoreau, F. & Weston, E. M. Evolving between land and water: key questions on the emergence and history of the Hippopotamidae (Hippopotamidae, Cetartiodactyla). *Biol. Rev. Camb. Phil. Soc.* **86**, 601–625 (2011).
- Bibi, F. Mio-pliocene faunal exchanges and African biogeography: the record of fossil bovids. *PLoS ONE* **6**, e16688 (2011).
- Zhang, Z. *et al.* Aridification of the Sahara desert caused by Tethys Sea shrinkage during the Late Miocene. *Nature* **513**, 401–404 (2014).
- Scally, A. *et al.* Insights into hominid evolution from the gorilla genome sequence. *Nature* **483**, 169–175 (2012).
- Hara, Y., Imanishi, T. & Satta, Y. Reconstructing the demographic history of the human lineage using whole-genome sequences from human and three great apes. *Genome Biol. Evol.* **4**, 1133–1145 (2012).
- Langergraber, K. E. *et al.* Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc. Natl Acad. Sci. USA* **109**, 15716–15721 (2012).
- Tsujikawa, H. The updated late Miocene large mammal fauna from Samburu Hills, northern Kenya. *Afr. Study Monogr.* **32** (suppl.), 1–50 (2005).
- Leakey, M. G. & Harris, J. M. *Lothagam: The Dawn of Humanity in Eastern Africa*. (Columbia Univ. Press, 2003).
- Boisserie, J.-R. *et al.* A new species of *Nyanzachoerus* (Cetartiodactyla: Suidae) from the late Miocene Toros-Ménalla, Chad, central Africa. *PLoS ONE* **9**, e103221 (2014).
- Chernet, T., Hart, W., Aronson, J. & Walter, R. C. New age constraints on the timing of volcanism and tectonism in the northern Main Ethiopian Rift–southern Afar transition zone (Ethiopia). *J. Volcanol. Geotherm. Res.* **80**, 267–280 (1998).
- Wolfenden, E., Ebinger, C., Yirgu, G., Deino, A. & Ayalew, D. Evolution of the northern Main Ethiopian rift: birth of a triple junction. *Earth Planet. Sci. Lett.* **224**, 213–228 (2004).
- Kunz, K., Kreuzer, H. & Müller, P. in *Afar Depression of Ethiopia* (eds Pilger, A. & Rössler, A.) 370–374 (Schweizerbart, 1975).
- Suwa, G. *et al.* Newly discovered 7 to 9 million-year-old cercopithecoid, equid and other mammalian fossils from the Chorora Formation, Ethiopia. *Anthropol. Sci.* **123**, 19–39 (2015).
- Benefit, B. R. & Pickford, M. Miocene fossil cercopithecoids from Kenya. *Am. J. Phys. Anthropol.* **69**, 441–464 (1986).
- Gilbert, C. C., Goble, E. D. & Hill, A. Miocene cercopithecoidea from the Tugen Hills, Kenya. *J. Hum. Evol.* **59**, 465–483 (2010).
- Bernor, R. L., Armour-Chelu, M. J., Gilbert, H., Kaiser, T. M. & Schulz, E. in: *Cenozoic Mammals of Africa* (eds Werdelin, L. & Sanders, W. J.) 685–721 (Univ. California Press, 2010).
- Uno, K. T. *et al.* Late Miocene to Pliocene carbon isotope record of differential diet change among East African herbivores. *Proc. Natl Acad. Sci. USA* **108**, 6509–6514 (2011).
- Eronen, J. T. *et al.* Distribution history and climatic controls of the Late Miocene Pikermian chronofauna. *Proc. Natl Acad. Sci. USA* **106**, 11867–11871 (2009).
- Flynn, L. J. *et al.* The Leporid Datum: a late Miocene biotic marker. *Mammal Rev.* **44**, 164–176 (2013).
- Hobolth, A., Dutheil, J. Y., Hawks, J., Schierup, M. H. & Mailund, T. Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. *Genome Res.* **21**, 349–356 (2011).
- Begun, D. R., Nargolwalla, M. C. & Kordos, L. European Miocene hominids and the origin of the African ape and human clade. *Evol. Anthropol.* **21**, 10–23 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank the Authority for Research and Conservation of Cultural Heritage, Ministry of Culture and Tourism of Ethiopia, for permissions and facilitation; we thank the Western Hararge Chiro Zone Culture and Tourism Office and the Mieso Woreda for fieldwork support; we thank all participants in the fieldwork, especially the Gololcha people, who were essential to the success of the project. Neutron irradiation for  $^{40}\text{Ar}$ – $^{39}\text{Ar}$  dating was done under the Visiting Researchers Program at the Kyoto University Research Reactor Institute. We thank J. Morton for assistance with major and trace element analyses, and the Janet and Elliott Banes Professorship and National Science Foundation EAR-1028789 for support to W.H. We thank H. Ishiguro for assistance with sampling enamel for the isotope analysis, and M. Nakatsukasa and Y. Kunimatsu for comparative materials. We thank D. Geraads for providing field information about previous geochronological sampling localities. This project was supported primarily by the Japan Society for the Promotion of Science (Kakenhi grant numbers 21255005 and 24000015).

**Author Contributions** G.S., Y.B., B.A., S.K. designed the research; G.S., Y.B., B.A., S.K., T.S., K.S. conducted field work; S.K., T.I., H.H., M.H., K.Y., C.G., G.W., W.H. did the geochronological and geochemical analysis; G.S., H.N., R.B., J.-R.B., F.B., H.S., S.A. did the faunal and isotopic analysis; and G.S., Y.B., B.A., S.K. wrote the manuscript with contributions from all co-authors.

**Author Information** The Chorora Formation vertebrate fossils have been deposited in the Paleontology and Paleoanthropology laboratories of the Authority for Research and Conservation of Cultural Heritage, Ministry of Culture and Tourism, Addis Ababa, Ethiopia. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.S. ([gusuwa@um.u-tokyo.ac.jp](mailto:gusuwa@um.u-tokyo.ac.jp)).

## METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

**Lithostratigraphic and tephrostratigraphic correlations.** The Chorora Formation sediments and volcanic rocks were mapped and described in the post-2010 field surveys at the following localities: type locality, Beticha, Gololcha, Chifara, Adadi, Gutosadeen, Odakora North and Teso Tadecho. Synthesized lithostratigraphic columns were constructed for each locality and correlations between localities were established (Extended Data Figs 5 and 6). The type locality and Beticha sediments are dominated by diatomaceous clay, silt and intercalated tuffs. At Gololcha and the lower Chifara sections, coarser tuffs and pyroclastic flows dominate, with intercalated partly diatomaceous silt and sand beds. These were compared with previous depictions of the type locality and nearby stratigraphy<sup>2,31</sup>. At Adadi and Gutosadeen, stratigraphically higher sediments unconformably overlie the sediments correlative to the type locality and Beticha lower beds.

Tephrostratigraphy of the Chorora Formation tuffs was compiled on the basis of lithology (thickness, colour, grain size, unit division), stratigraphic relationships and petrographic properties (including morphology and refractive indices of volcanic glass shards). Probable correlative tephra layers were established by the analytical results (Supplementary Table 1). Eight tuffs (CT-1 to CT-8) were confirmed to be correlative by chemical fingerprinting of discrete volcanic glass shards (Supplementary Tables 1 and 2). Major element compositions were measured by an electron-probe micro-analyzer and processed following established methods<sup>32,33</sup>. The coefficient of similarity values<sup>34</sup> were determined using eight major element oxides: SiO<sub>2</sub>, TiO<sub>2</sub>, Al<sub>2</sub>O<sub>3</sub>, FeO, MnO, CaO, Na<sub>2</sub>O and K<sub>2</sub>O. All coefficient of similarity values calculated for correlative pairs exceeded 0.91 (Supplementary Table 1). For two of the tephra (CT-5 and CT-6), after examining chemical homogeneity of glass shards, trace element compositions of glass shards were measured by direct current plasma optical emission spectrometry (DCP-OES) following established methods<sup>33,35</sup> (Supplementary Table 2). For three of the tephra (CT-4, CT-7 and CT-8), 10 major and 21 trace elements were also analysed by inductively coupled plasma optical emission spectrometry (ICP-OES) (Agilent 720ES ICP-OES by W. K. Hart and J. P. Morton). One sample of tuff CT-5 (07T-2) was analysed in 2010 (DCP-OES) and again in 2014 (ICP-OES), along with in-house glass standards and an international standard reference material, to establish method/instrument comparability and overall analytical accuracy and precision (Supplementary Table 2).

The ICP-OES techniques were similar to those previously used for DCP-OES analyses with the following important distinctions. Major and trace elements were analysed simultaneously using a single dissolution of 50 mg of sample powder fused with 75 or 100 mg of lithium metaborate and dissolved in 125 ml of 1% nitric acid. The resulting solution had a low enough amount of total dissolved solids to be run without further dilution. Powders of 12–14 standard reference materials were used as external calibration standards, and were chosen to emphasize the range of compositions of the unknown materials. These were prepared the same way and fused at the same time as the unknowns. Samples and standards were analysed within 24 h after fusion to minimize losses of Si and Zr.

**K–Ar dating.** Fresh volcanic rocks (ignimbrites, lavas and tuffs) were collected for K–Ar and <sup>40</sup>Ar–<sup>39</sup>Ar dating. Sample preparation and K–Ar dating was conducted following established methods<sup>36,37</sup>. Feldspar separates of 60–100, 100–200, 150–200, 150–300 or 200–300 mesh size fractions (250–150 µm, 150–74 µm, 105–74 µm, 105–48 µm and 74–48 µm, respectively) were prepared for analysis and examined to assess chemical homogeneity by electron-probe microanalysis following established methods<sup>37</sup>. Seven major element oxides, SiO<sub>2</sub>, TiO<sub>2</sub>, Al<sub>2</sub>O<sub>3</sub>, FeO, CaO, Na<sub>2</sub>O and K<sub>2</sub>O, and two minor oxides, SrO and BaO, were analysed using at least 25 crystals for each sample. The results are shown in the Or–Ab–An (K–Na–Ca) diagrams (Extended Data Fig. 3) and in Supplementary Table 3; the crystals analysed were predominantly alkali-feldspars (orthoclase and anorthoclase) and minor plagioclases in the felsic units, or plagioclase in the basalt lava samples. A higher chemical homogeneity and finer-grained separates generally result in higher accuracy of K–Ar ages, so that 100–200 or 150–200 mesh size fractions were preferentially used, except for the basalt lava samples for which 150–300 or 200–300 mesh size fractions were dated to further minimize the heterogeneity effect.

**<sup>40</sup>Ar–<sup>39</sup>Ar dating.** Single crystal step-heating <sup>40</sup>Ar–<sup>39</sup>Ar ages of selected samples were measured for comparisons with the K–Ar ages. Fresh coarse-grained anorthoclase crystals (0.5–1.5 mm in size) were taken from the feldspar separate prepared for the K–Ar dating. Ages of individual crystals were determined by stepwise degassing using a near-infrared diode laser<sup>8,38</sup>. Plateau ages were calculated from fractions that spanned over 75% of the total released <sup>39</sup>Ar and overlapped each other within 2σ error<sup>39</sup>, except for two samples (10TC-2 and 10TC-42) in which higher calcium phases caused age shifts.

At Beticha, units above and below the *Chororapithecus* fossil-bearing sands were dated. Two samples of the overlying consolidated tuff were taken ~5 m apart at

(10TC-58) or near (10TC-1) the cliff face just north of the *Chororapithecus* discovery area. From these two samples, a total of six individual crystals were analysed (Extended Data Fig. 4). The step-heating single crystal ages ranged from 7.59 ± 0.24 (2σ) to 8.19 ± 0.26 Ma with a weighted mean age of 7.82 ± 0.11 Ma. Two samples of the underlying pumiceous tuff (10TC-52 and 10TC-53) were taken ~30 m apart just downslope and down section of the *Chororapithecus* fossil-bearing unit. From the two samples, a total of six individual crystals were analysed (Extended Data Fig. 4). The step-heating single crystal ages ranged from 7.70 ± 0.28 to 7.93 ± 0.24 Ma with a weighted mean age of 7.86 ± 0.10 Ma. We also determined the step-heating single crystal <sup>40</sup>Ar–<sup>39</sup>Ar ages of the rhyolitic ignimbrite at Beticha (10TC-02) and Gololcha (10TC-42). Two feldspar crystals were measured for each sample, yielding plateau ages of one crystal each (10.20 ± 0.48 and 9.52 ± 0.24 Ma, respectively).

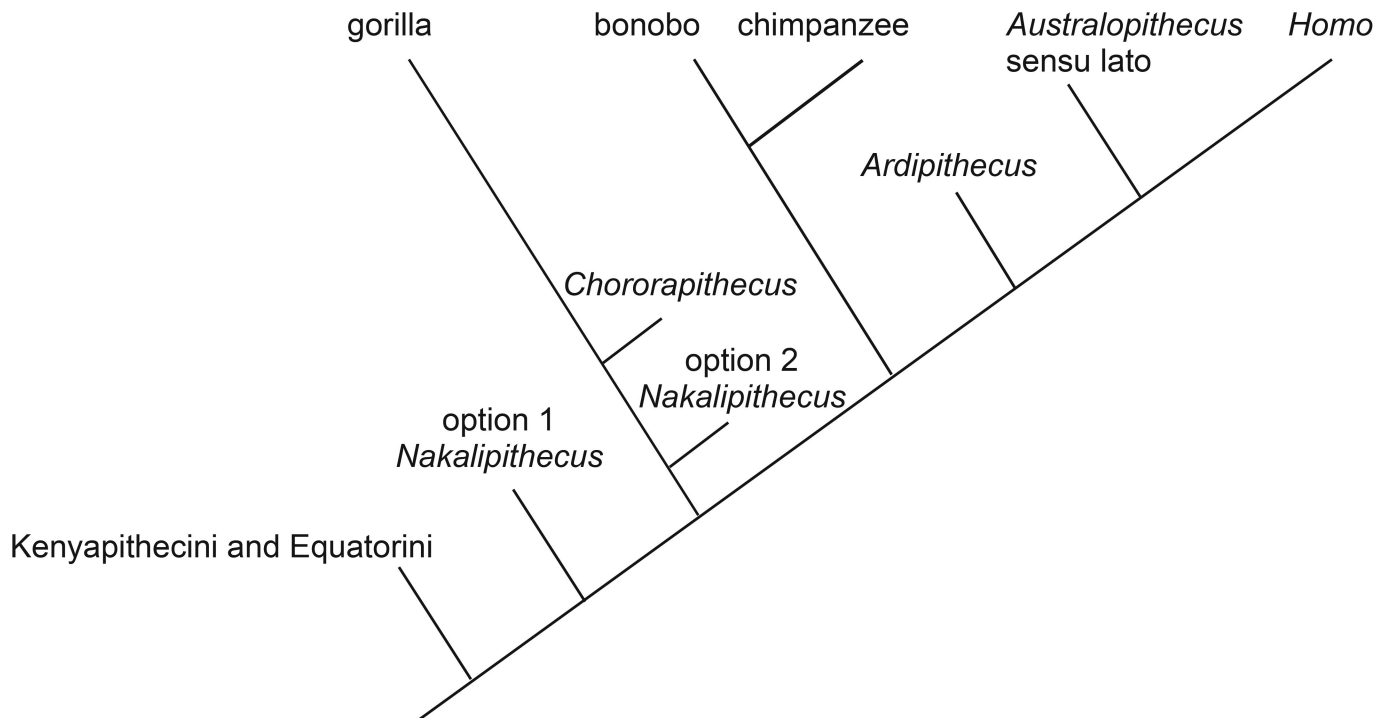
**Magnetostratigraphy.** Oriented block samples were taken from the Chorora Formation sediments and volcanic rocks following established methods<sup>40</sup>. Cubic subsamples were cut from weakly consolidated sediments with stainless and aluminium saws and sand papers and put into cubic holders of 10 cm<sup>3</sup>. Columnar subsamples of about 10 cm<sup>3</sup> were mechanically cored from the hard rock samples. Stepwise alternating field and thermal demagnetizations were applied to these subsamples in 15–17 steps (2.5–180 mT and 100–680 °C), using a 2G cryogenic magnetometer and a Barington MS2 susceptibility meter. After confirming the natural remanent magnetizations with a linear decay towards the origin, characteristic remanent magnetization directions were calculated using principle component analysis<sup>41</sup>. A maximum angle deviation of 11.0° for at least five successive steps was accepted as the criterion in determining characteristic remanent magnetization direction. Mean palaeomagnetic directions were calculated for each sample horizon on the basis of the characteristic remanent magnetization directions for declination, inclination, precision parameter (*k*) and radius of 95% confidence circle (α<sub>95</sub>) as well as virtual geomagnetic pole latitude and longitude (Supplementary Table 4). The mean palaeomagnetic directions were less than 20° of α<sub>95</sub> in most of the sample horizons. The results indicate an upward normal–reverse–normal polarity change within the Chorora Formation sediments ('series 1' and 'series 2' sediments of Fig. 3) and a reversed polarity of the underlying rhyolitic ignimbrite. The primate-fossil-bearing horizon at Beticha occurs just above the transition from reverse to normal polarity. Considering our K–Ar and <sup>40</sup>Ar–<sup>39</sup>Ar ages, we interpret this transition as correlative to the boundary between C4r and C4n.2n, 8.072 Ma according to the Cande and Kent geomagnetic polarity timescale<sup>42</sup>.

**Isotopic analysis of enamel.** The results of carbon and oxygen stable isotope analysis of equid and hippopotamid tooth enamel are summarized in Extended Data Table 3, together with the enamel δ<sup>13</sup>C values of four previously reported hipparionin teeth<sup>422</sup>. We measured enamel δ<sup>13</sup>C and δ<sup>18</sup>O values on three of the newly collected hipparionin teeth from the type locality (~8.5 Ma) and two hippopotamid fragments from Beticha (~8.0 Ma) following established methods<sup>43</sup>. Diagenetic carbonate was removed from powdered enamel by treatment for 4 h with 0.1 M acetic acid. Results are expressed as parts per thousand (‰) difference (δ) from the <sup>13</sup>C/<sup>12</sup>C and <sup>18</sup>O/<sup>16</sup>O ratios of international standards (Pee Dee Belemnite (PDB) fossil carbonate for carbon; PDB and Vienna Standard Mean Ocean Water for oxygen).

Accurate estimation of past atmospheric CO<sub>2</sub> δ<sup>13</sup>C values is critical for reconstructing Late Miocene habitats and herbivore diets. The modern C<sub>3</sub> global average δ<sup>13</sup>C value, excluding closed canopy forests, is approximately –27‰ (ref. 44), and the pre-industrial average would have been –25.5‰. Estimates of Neogene air δ<sup>13</sup>C values based on marine carbonate and organic biomarker isotopic data are similar to preindustrial values<sup>45,46</sup>. Air CO<sub>2</sub> concentrations are higher and δ<sup>13</sup>C values are lower in closed canopy forest floor microenvironments. Discrimination (Δ<sup>13</sup>C‰ = air δ<sup>13</sup>C‰ – plant δ<sup>13</sup>C‰) varies substantially with temperature, humidity, water stress, CO<sub>2</sub> concentration and functional type (evergreen gymnosperms discriminate less than deciduous angiosperms), ranging from 22.5‰ in evergreen warm mixed forests to 15.6‰ in xeric woodland/scrubland<sup>47</sup>. Tropical habitats with mean annual rainfall ~500 to 1,500 mm have estimated average Δ<sup>13</sup>C values of ~19 to 22‰ (ref. 47). High CO<sub>2</sub> concentrations during the Palaeogene and early Neogene favoured C<sub>3</sub> plants, but decreasing concentrations led to the expansion of C<sub>4</sub> grasslands during the Late Miocene, beginning as early as ~10 Ma. Modern browsing species generally do not feed on gymnosperm leaves, so we assume that browsing herbivores consumed plants with higher Δ<sup>13</sup>C values, and thus more negative δ<sup>13</sup>C than the total biome average. If the Late Miocene atmospheric CO<sub>2</sub> average δ<sup>13</sup>C value was –6.5‰, and precipitation was 500–1,500 mm, then the average angiosperm plant Δ<sup>13</sup>C value may have been ~20‰, so the average Late Miocene C<sub>3</sub> δ<sup>13</sup>C values would have been –26.5‰. Isotopic enrichment from ruminant herbivore diet to tooth enamel carbonate of ~14‰ would result in δ<sup>13</sup>C values of approximately –12.5‰ and 2–3‰ for pure C<sub>3</sub> feeders (browsers) and pure C<sub>4</sub> feeders (grazers), respectively. Values between these end-members reflect linear proportions of C<sub>3</sub>- and C<sub>4</sub>-based foods. The canopy effect can shift C<sub>3</sub> browser enamel δ<sup>13</sup>C values to as low as –20‰ in the Congo forest<sup>48</sup>.

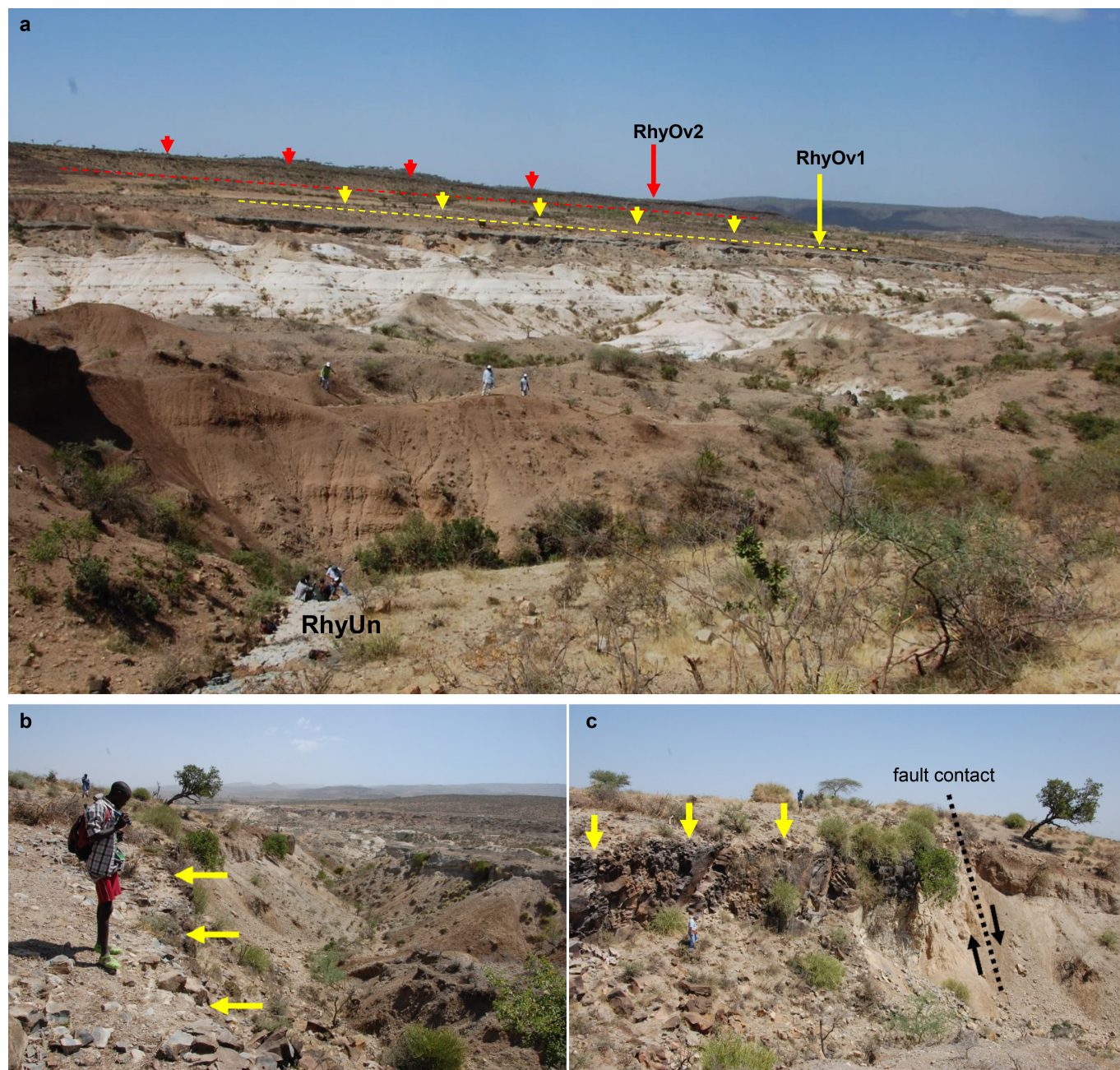
31. Tiercelin, J. J. & Michaux, J. Un exemple de remplissage mio-pliocène de rift dans l'Est africain: la Formation de Ch'orora, Éthiopie: stratigraphie et mammifères fossils. In *5<sup>e</sup> Réunion. Sc. Terre, Renne*, **449** (1977).
32. Katoh, S. *et al.* Chronostratigraphy and correlation of the Plio-Pleistocene tephra layers of the Konso Formation, southern main Ethiopia rift, Ethiopia. *Quat. Sci. Rev.* **19**, 1305–1317 (2000).
33. WoldeGabriel, G., Hart, W. K., Katoh, S., Beyene, Y. & Suwa, G. Correlation of Plio-Pleistocene tephra in Ethiopian and Kenyan rift basins: temporal calibration of geological features and hominid fossil records. *J. Volcanol. Geotherm. Res.* **147**, 81–108 (2005).
34. Borchardt, G. A., Harward, M. E. & Schmitt, R. A. Correlation of volcanic ash deposits by activation analysis of glass separates. *Quat. Res.* **1**, 247–260 (1971).
35. Katoh, S., Danhara, T., Hart, W. K. & WoldeGabriel, G. Use of sodium polytungstate solution in the purification of volcanic glass shards for bulk chemical analysis. *Nat. Hum. Activities* **4**, 45–54 (1999).
36. Itaya, T. *et al.* Argon analysis by a newly-developed mass spectrometric system for K-Ar dating. *Mineral. J.* **15**, 203–221 (1991).
37. Ryu, S., Oka, M., Yagi, K. & Itaya, T. K-Ar ages of the Quaternary basalts in the Jeongok area, the central part of Korean peninsula. *GeoJournal* **15**, 1–8 (2011).
38. Hyodo, H., Kim, S.-W., Itaya, T. & Matsuda, T. Homogeneity of neutron flux during irradiation for  $^{40}\text{Ar}/^{39}\text{Ar}$  age dating in the research reactor at Kyoto University. *J. Min. Petr. Econ. Geol.* **94**, 329–337 (1999).
39. Kawamura, M., Uchino, T., Gouzu, C. & Hyodo, H. 380 Ma  $^{40}\text{Ar}/^{39}\text{Ar}$  ages of the high-P/T schists obtained from the Nedamo Terrane, Northeast Japan. *J. Geol. Soc. Japan* **113**, 492–499 (2007).
40. Hyodo, M. *et al.* High-resolution record of the Matuyama-Brunhes transition constrains the age of Javanese *Homo erectus* in the Sangiran dome, Indonesia. *Proc. Natl Acad. Sci. USA* **108**, 19563–19568 (2011).
41. Kirschvink, J. L. The least-squares line and plane and the analysis of palaeomagnetic data. *Geophys. J. R. Astron. Soc.* **62**, 699–718 (1980).
42. Cande, S. C. & Kent, D. V. Revised calibration of the geomagnetic polarity time scale for the Late Cretaceous and Cenozoic. *J. Geophys. Res.* **100**, 6093–6095 (1995).
43. White, T. D. *et al.* Macrovertebrate paleontology and the Pliocene habitat of *Ardipithecus ramidus*. *Science* **326**, 67, 87–93 (2009).
44. Kohn, M. J. Carbon isotope compositions of terrestrial  $\text{C}_3$  plants as indicators of (paleo)ecology and (paleo)climate. *Proc. Natl Acad. Sci. USA* **107**, 19691–19695 (2010).
45. Tipple, B. J., Meyers, S. R. & Pagani, M. Carbon isotope ratio of Cenozoic  $\text{CO}_2$ : a comparative evaluation of available geochemical proxies. *Paleoceanography* **25**, PA3202 (2010).
46. Feakins, J. J. *et al.* Northeast African vegetation change over 12 m.y. *Geology* **41**, 295–298 (2013).
47. Diefendorf, A. F., Mueller, K. E., Wing, S. L., Koch, P. L. & Freeman, K. H. Global patterns in leaf  $^{13}\text{C}$  discrimination and implications for studies of past and future climate. *Proc. Natl Acad. Sci. USA* **107**, 5738–5743 (2010).
48. Cerling, T. E., Harris, J. M., Leakey, M. G. & Mudida, N. in *Lothagam: The Dawn of Humanity in Eastern Africa* (eds Leakey, M. G., Harris, J. M.) 584–603 (Columbia Univ. Press, 2003).
49. Boissérie, J.-R. *et al.* Diets of modern and late Miocene hippopotamids: evidence from carbon isotope composition and micro-wear of tooth enamel. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **221**, 153–174 (2005).





**Extended Data Figure 1 | Cladistic relationships of *C. abyssinicus* and *N. nakayamai*.** *C. abyssinicus* is known from only nine teeth (or fragments)<sup>6</sup> and *N. nakayamai* from a mandibular corpus with M1–M3, and ten other isolated teeth (excluding a possible antimer)<sup>8</sup>. The only informative dental elements shared by the two reported samples are the lower M3 and the lower M1 (the latter damaged in *C. abyssinicus* and considerably worn in *N. nakayamai*). This fragmentary evidence makes cladistic evaluations difficult. *C. abyssinicus* was considered to share the following derived combination of features with the modern gorilla: large postcanine size, upper molars buccolingually narrow and mesiodistally elongate, with a relatively long and mesiobuccally extending mesial protocone crest, reduced protoconule, and lower molars with a prominent anterior transverse crest. Following ref. 6, we consider *Chororapithecus* to be a basal member of the gorilla clade. When compared with the Middle Miocene examples of Kenyapithecini and Equatorini, *C. abyssinicus* and *N. nakayamai* share reduced cingula in both upper and lower molars,

although ref. 8 pointed out that the lower M3 cingulum appears slightly better developed in *N. nakayamai*. The available *N. nakayamai* teeth include elements (not reported in *C. abyssinicus*) that are morphologically more derived than in Kenyapithecini/Equatorini, such as upper premolars that are relatively elongate mesiodistally (narrow buccolingually) and a lower P3 that is not as obliquely elongate and transversely compressed as in the Middle Miocene forms. On the basis of these observations, ref. 8 preferred a stem modern African ape–human cladistic position of *N. nakayamai* (option 1). However, *N. nakayamai* also shares two possibly derived features with *C. abyssinicus*: large postcanine size (although slightly less so) and prominent transverse crest in the lower molars. The latter is inferred from the characteristic wear pattern of the holotype mandible lower M1, the dentine exposures of the lingual cusps that are buccolingually linear (figure 1 of ref. 8). These observations suggest that *N. nakayamai* may also be a stem member of the gorilla clade but more primitive than *C. abyssinicus* (option 2).

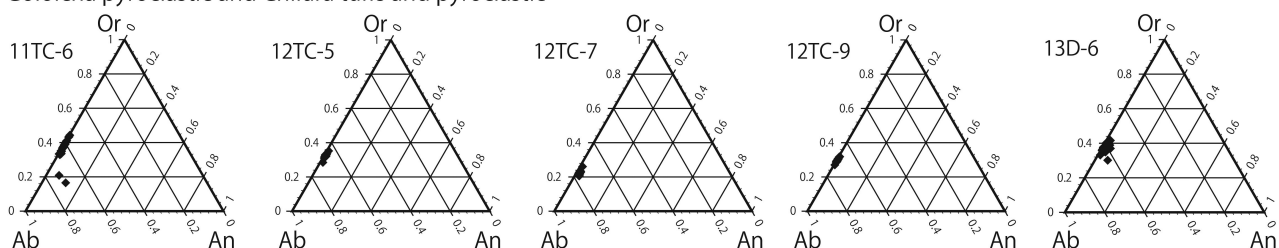
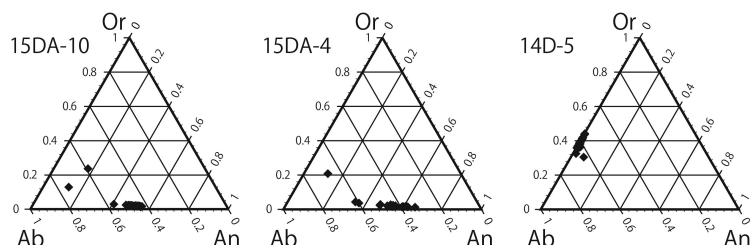
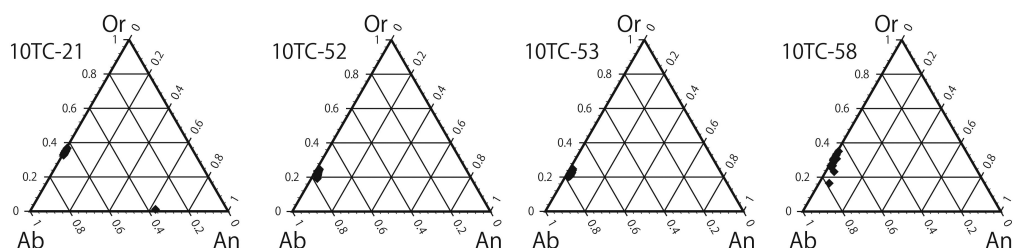


**Extended Data Figure 2 | Chorora Formation type locality.** **a**, View north-northwest from the southeastern margin of the type locality. Yellow and red arrows show the rhyolitic ignimbrite units considered by previous workers to overlie the Chorora Formation sediments<sup>1–3</sup>. The ignimbrite unit (yellow arrows) adjacent to the sediment exposures has been interpreted to directly overlie the Chorora Formation sediments including the gravel unit that forms the resistant ledge along the western margin of the whitish lacustrine exposures. A second ridge top unit (red arrows) was considered to cap the entire sequence. RhyUn, RhyOv1 and RhyOv2 indicate sampling localities of the present study for K–Ar dating. The

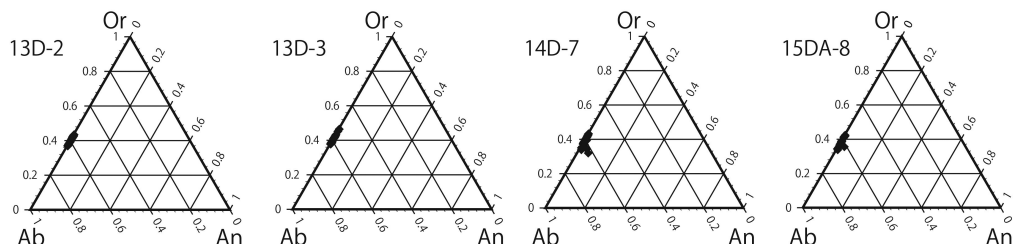
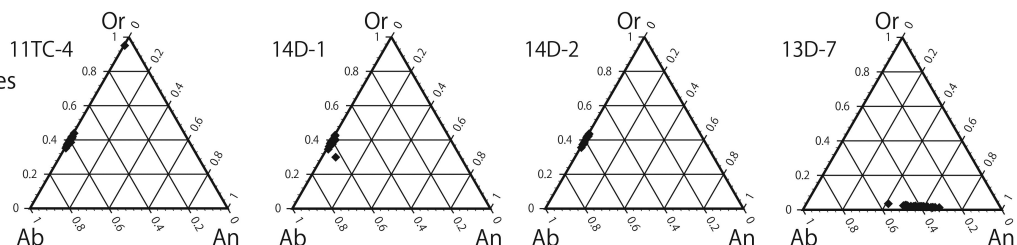
RhyUn unit dips westwards and stratigraphically underlies the Chorora Formation sediments. The horizontally bedded light brown sediments are Middle Pleistocene in age and unconformably cap both RhyUn and Chorora Formation sediments. The yellow and red dotted lines indicate approximate positions of the step faults that resulted in uplifted RhyOv1 and RhyOv2 exposures. **b**, View northeast at the RhyOv1 sampling spot. Yellow arrows point to the RhyOv1 rhyolitic ignimbrite unit forming a planar terrace. **c**, View west of the **b** RhyOv1 section (yellow arrows). Note that the rhyolitic ignimbrite unit terminates abruptly (dotted line) in fault contact with sediments previously considered to underlie the ignimbrite<sup>2,3</sup>.



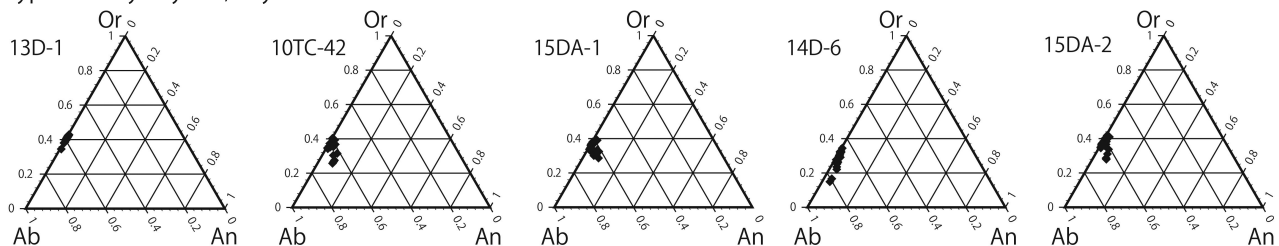
## Gololcha pyroclastic and Chifara tuffs and pyroclastic

Gutosadeen and Teso Tadecho basalt  
Odakora North ignimbriteType Locality  
and Beticha tuffs

## Type Locality RhyUn

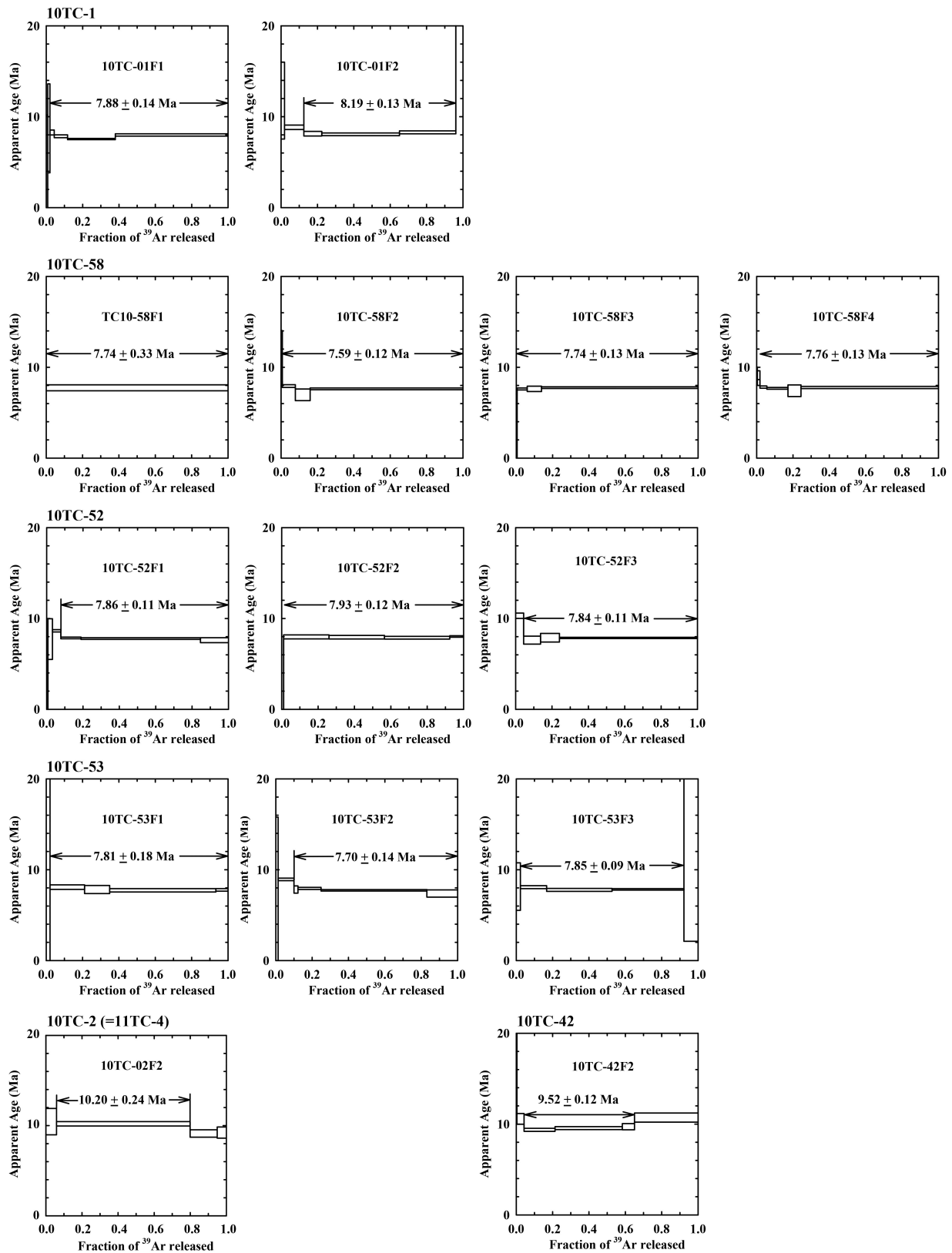
Beticha RhyUn correlatives  
and basalt

## Type Locality RhyOv1, RhyOv2 and Gololcha correlatives



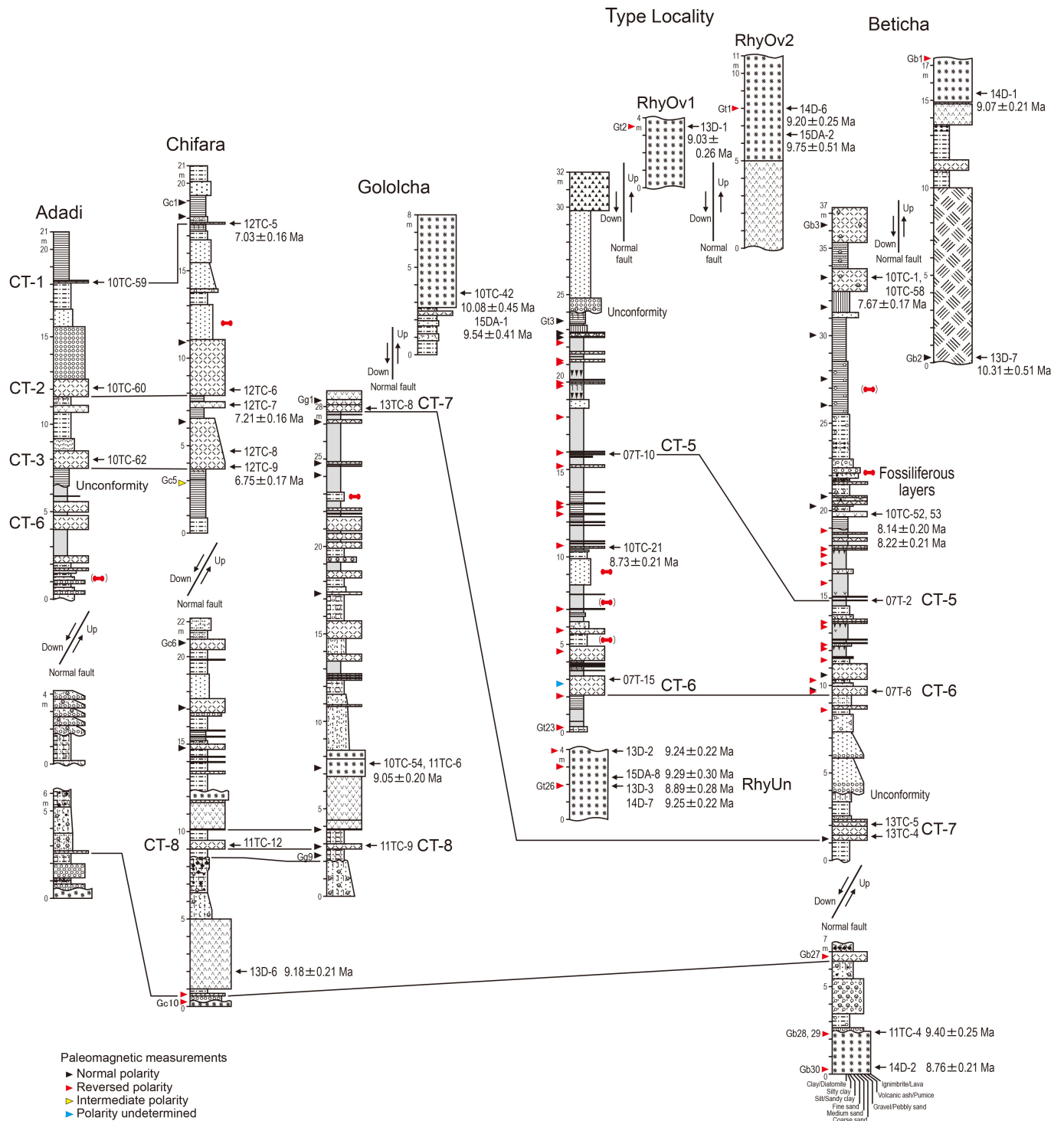
**Extended Data Figure 3 | Electron-probe microanalysis (Or-Ab-An diagrams) of the K-Ar-dated samples.** Results of the electron-probe microanalysis of feldspar separates of the mesh size fraction that was used in the K-Ar dating of samples listed in Extended Data Table 1. The full data are shown in Supplementary Table 3.





**Extended Data Figure 4 | Single crystal step-heating  $^{40}\text{Ar}$ - $^{39}\text{Ar}$  analysis.** Results of individual crystals with stable ages (indicated by arrows) are shown. Errors are  $1\sigma$ . Samples 10TC-1 and 10TC-58 are from the same

consolidated tuff unit collected <5 m apart. Samples 10TC-52 and 10TC-53 are from the same pumiceous tuff collected ~30 m apart. The weighted means of these are shown in Fig. 3 and Extended Data Table 1.



**Extended Data Figure 5 | Chorora Formation stratigraphic columns of the type locality, Beticha and nearby localities.** The sections were taken at exposures adjacent to or continuous with the sampling locations of the dated volcanic samples (coordinates tabulated in Extended Data Table 1). Radioisotopic dates (K–Ar ages) and samples correspond to those shown in Fig. 3 and Extended Data Table 1. Eight tephra units are

considered to occur at multiple localities, and shown as marker tuffs CT-1 to CT-8. The analytical details of the tuffs are presented in Supplementary Tables 1 and 2. CT-6 is identified at Adadi from lithologic, petrographic, glass morphology and refractive index analyses. Results of the remanent magnetism analysis are given in Supplementary Table 4. See Extended Data Fig. 6 for explanation of the lithological codes.





Extended Data Table 1 | Summary of K–Ar and  $^{40}\text{Ar}$ – $^{39}\text{Ar}$  dating

No.	Sample	Locality	Stratigraphic position	Material	Mesh size	K wt%	Rad $^{40}\text{Ar}$ $10^{-8}\text{cc}$ STP/g	Nonrad $^{40}\text{Ar}\%$	K–Ar age (Ma)	$2\sigma$	$^{40}\text{Ar}/^{39}\text{Ar}$ age (Ma)	$2\sigma$
1	12TC-5 N 8.8817, E 40.3042	Chifara	above "series2"	Pm-tuff	#60–100	4.760±0.095	130.1±1.3	2.9	7.03	0.16		
2	12TC-7 N 8.8818, E 40.3052	Chifara	above "series2"	Tuff	#60–100	3.224±0.064	90.4±0.9	6.0	7.21	0.16		
3	12TC-9 N 8.8817, E 40.3052	Chifara	above "series2"	Tuff	#60–100	4.443±0.089	116.6±1.7	28.9	6.75	0.17		
4	15DA-4 N 8.8589, E 40.2548	Teso Tadecho	above "series2"	Basalt	#200–300	1.243±0.025	32.49±0.45	23.9	6.72	0.16		
5	15DA-10 N 8.9043, E 40.2930	Gutosadeen	above "series 2"	Basalt	#200–300	1.318±0.026	35.59±0.53	31.1	6.94	0.17		
6	14D-5 N 8.8757, E 40.2726	Odakora North	above "series 2"	RI	#100–200	3.927±0.079	115.7±2.2	42.5	7.57	0.21		
7	10TC-1/58 N 8.8660, E 40.2948	Beticha	upper "series 2"	Pm-tuff	#100–200	3.905±0.078	116.5±1.3	9.5	7.67	0.17	7.82 (n=6)	0.11
8a	10TC-52/53	Beticha	upper "series 2"	Pm-tuff	#150–200	3.218±0.064	101.9±1.5	30.8	8.14	0.20	7.86 (n=6)	0.10
8b		same as above			#150–200	3.200±0.064	102.3±1.7	36.5	8.22	0.21		
	N 8.8655, E 40.2949 (10TC-52) N 8.8654, E 40.2946 (10TC-53)											
9	10TC-21b N 8.8900, E 40.3068	Type	middle "series 2"	Pm-tuff	#150–200	4.115±0.082	139.7±1.7	19.7	8.73	0.21		
10	10TC-54/11TC-6 N 8.8999, E 40.3120	Gololcha	middle "series 1"	Pyr	#150–200	4.075±0.082	143.4±1.5	10.7	9.05	0.20		
11	13D-6 N 8.8806, E 40.3047	Chifara	lower "series 1"	Pyr	#100–200	5.032±0.101	179.7±1.9	5.7	9.18	0.21		
12	13D-2	Type†	RhyUn	RI	#100–200	5.469±0.109	196.6±2.6	22.7	9.24	0.22		
13	13D-3	Type†	RhyUn	RI	#100–200	5.276±0.106	182.4±4.5	41.9	8.89	0.28		
14	14D-7	Type†	RhyUn	RI	#100–200	5.351±0.107	192.5±2.4	21.2	9.25	0.22		
15	15DA-8 N 8.8883, E 40.3071 (13D-2, 13D-3, 14D-7, 15DA-8)	Type†	RhyUn	RI	#100–200	5.024±0.100	181.7±4.5	53.4	9.29	0.30		
16	13D-1 N 8.8940, E 40.3074	Type	RhyOv1	RI	#100–200	5.461±0.109	191.9±4.1	46.0	9.03	0.26		
17	10TC-42 N 8.9014, E 40.3122	Gololcha	RhyOv1	RI	#100–200	2.073±0.041	81.3±3.2	68.5	10.08	0.45	9.52	0.22
18	15DA-1 N 8.9015, E 40.3123	Gololcha	RhyOv1	RI	#100–200	4.511±0.090	167.4±6.4	72.9	9.54	0.41		
19	14D-6 N 8.8940, E 40.3042	Type	RhyOv2	RI	#100–200	4.994±0.100	178.8±3.4	42.0	9.20	0.25		
20	15DA-2 N 8.8892, E 40.3012	Type	RhyOv2	RI	#100–200	4.580±0.092	173.8±8.4	72.9	9.75	0.51		
21	10TC-2/11TC-4	Beticha‡	RhyUn correlative	RI	#100–200	3.800±0.076	138.9±2.4	38.0	9.40	0.25	10.20	0.48
22	14D-2 N 8.8649, E 40.2953 (10TC-2/11TC-4, 14D-2)	Beticha‡	RhyUn correlative	RI	#100–200	5.367±0.107	182.9±2.3	19.9	8.76	0.21		
23	14D-1 N 8.8659, E 40.2929	Beticha	RhyUn correlative	RI	#100–200	5.455±0.109	192.6±2.1	12.0	9.07	0.21		
24	13D-7 N 8.8651, E 40.2928	Beticha	below 14D-1	Basalt	#150–300	0.370±0.007	14.85±0.7	71.3	10.31	0.51		
K–Ar age weighted means												
Nos. 8a–8b		Beticha	upper "series 2"	Pm-tuff (n=2)					8.18	0.15		
Nos. 12–15		Type	RhyUn	RI (n=4)					9.18	0.12		
Nos. 16–18		Type/Gololcha	RhyOv1	RI (n=3)					9.36	0.20		
Nos. 19–20		Type	RhyOv2	RI (n=2)					9.31	0.23		
Nos. 21–23		Beticha	RhyUn correlative	RI (n=3)					9.04	0.13		
Nos. 12–20		Type/Gololcha	RhyUn/Ov1/Ov2	RI (n=9)					9.25	0.09		
Nos. 12–23		Type/Gololcha/Beticha	RhyUn/Ov1/Ov2	RI (n=12)					9.17	0.08		

K–Ar and single crystal step-heating  $^{40}\text{Ar}$ – $^{39}\text{Ar}$  ages are in million years; weighted means are shown with number of crystals or samples in parentheses. RI, rhyolitic ignimbrite; Pm-tuff, pumiceous tuff; Pyr, rhyolitic pyroclastics. RhyUn refers to the rhyolitic ignimbrite unit that stratigraphically underlies the Chorora Formation sediments at the type locality. RhyOv1 and RhyOv2 refer to type locality rhyolitic ignimbrite units that were previously considered to overlie the Chorora Formation sediments (see Fig. 2 and Extended Data Fig. 2).

†Samples 13D-2, 13D-3, 14D-7 and 15DA-8 were taken at the same RhyUn section (shown in Extended Data Fig. 2); 13D-2 was taken 2 m above the other three samples; the last three were taken at near-identical positions.

‡Samples 10TC-2/11TC-4 and 14D-2 were taken ~5 m from each other.

**Extended Data Table 2 | List of mammalian taxa recovered from the Chorora Formation, and comparison with the 9.5–9.9 Myr old Nakali/Samburu Hills fauna and the 6.5–7.4 Myr old Lothagam Lower Nawata Member fauna**

Nakali and Samburu Hills	Chorora	Type Locality	Beticha	Upper Chorora	Lothagam Lower+Upper Nawata
Primates					
<i>Samburupithecus kiptalami</i> (S)					
<i>Nakalipithecus nakayami</i> (N)	<i>Chororapithecus abyssinicus</i>		○		Hominidae sp. (U)
<i>Microcolobus</i> sp. (N)	cf. <i>Microcolobus</i>	○			
	Colobinae sp.		○	cf ○	Colobinae sp. A, sp. B (L+U)
	Cercopithecinae sp.		○	○large	<i>Parapapio lothagamensis</i> (L+U)
Cetartiodactyla					
<i>Kenyapotamus coryndoni</i> (N+S)	cf. <i>Kenyapotamus</i>	○			
	Hippopotaminae gen. et sp. nov.		○		
	Hippopotaminae sp.			○	<i>Archaeopotamus</i> spp.
<i>Nyanzachoerus</i> sp. large (N+S)	<i>Nyanzachoerus</i> sp.	○	○	cf ○	<i>Nyanzachoerus tulotos</i> (L+U)
? <i>Samotherium</i> sp.	Sivatheriini sp.	○	○		Sivatherium sp. (?L)
<i>Paleotragus</i> cf. <i>germaini</i> (N+S)	<i>Paleotragus</i> sp.	○	○		<i>Paleotragus germaini</i> (L+U)
Boselaphini sp. small (N?+S)	Boselaphini sp. small	○	○		Boselaphini sp. small (L+U)
	cf. Bovini		○		Bovini sp. (L+U)
? <i>Reduncini</i> (S)	<i>Reduncini</i> sp.	cf ○		○	<i>Reduncini</i> spp. (L+U)
	cf. Neotragini sp.	○			Neotragini spp. (L+U)
Perissodactyla					
“C.” aff. <i>africanum</i> (N+S)	“ <i>Cormhipparion</i> ” sp.	○	○	○?	<i>Eurygnathohippus turkanensis</i> (L+U)
small Hipparionini	Hipparionini sp. small		○	○	<i>Eurygnathohippus feibeli</i> (L+U)
<i>Paradicerus mukirii</i> (S)	<i>Ceratotherium</i> sp.	○			<i>Ceratotherium</i> sp. (L+U)
<i>Diceros</i> sp. (N)	cf. Dicerotini		○		<i>Diceros bicornis/praecox</i> (U)
small schizotheriinae	<i>Ancylotherium tugenensis</i>	○			
Proboscidea					
<i>Tetralophodon</i> sp nov. (S)					
Elephantidae sp. (N)	<i>Stegotetabelodon</i> sp.	○	○?	○	<i>Stegotetabelodon orbus</i> (L+U)
<i>Deinotherium</i> sp. (N+S)	<i>Deinotherium</i> sp.	○			<i>Deinotherium bozasi</i> (L+U)
Carnivora					
	? <i>Herpestides afarensis</i>	○			
Machairodontinae sp. (S)	<i>Machairodus</i> cf. <i>aphanistus</i>	○			<i>Lokotunjailurus emageritus</i> (L+U)
	Felidae sp.		large	medium	
<i>Percrocuta leakeyi</i> (N)	Percrocutidae sp. large		○		
Hyaenidae spp. (N+S)	cf. <i>Hyaenictis</i>	○			<i>Hyaenictis</i> sp. (L+U)
Lagomorpha					
	<i>Alilepus</i> sp.			○	<i>Alilepus</i> (L*)
Rodentia					
<i>Paraulacodus</i> sp. (N+S)	<i>Paraulacodus johanesi</i>	○	○	○	<i>Paraulacodus</i> sp. (L)
<i>Paraphiomys</i> sp. (N+S)	<i>Paraphiomys chororensis</i>	○			<i>Paraphiomys chororensis</i> (L)
<i>Nakalimys lavocati</i> (N)	<i>Nakalimys lavocati</i>	○			
Gerbillinae (N)	Gerbillinae sp. large cf. <i>Abudhabia</i>			○	<i>Abudhabia</i> sp. (U)
	<i>Afromys guillemoti</i>	○			
	“ <i>Dendromys</i> ” (? <i>Saccostomus</i> ) spp.	○			
	<i>Praecomys kikiae</i>	○			
	aff. <i>Stenocephalomys</i>	○			
	cf. <i>Parapelomys</i>	○			
	cf. <i>Tectonomys</i>	○			
	<i>Xerus</i> sp.	○			

Faunal list of Nakali and Samburu Hills adapted from refs 8, 16, and that of Lothagam from ref. 17. Only taxa that are related or phylogenetically close to the Chorora Formation taxa are listed. Grey shaded cells indicate taxa diagnostically distinct between the 9.5–9.9 Myr and 6.5–7.4 Myr old eastern African assemblages. Orange cells are Chorora Formation occurrences that suggest continuity with Nakali/Samburu Hills. Green cells are derived Chorora Formation taxa that indicate a biochronologically younger age than Samburu/Nakali and an older age than the Lothagam Lower Nawata Member. Blue cells indicate affinities between the Chorora Formation and Lower Nawata Member taxa.

\*Lower Nawata Member *Alilepus* at Lothagam comes from the 6.5–6.6 Myr old levels.

**Extended Data Table 3 | Values of  $\delta^{13}\text{C}$  and  $\delta^{18}\text{O}$  of Chorora Formation hipparionin and hippopotamid tooth enamel**

Specimen no.	Taxon	Lab Sample no.	Element	$\delta^{13}\text{C}\text{‰ PDB}$	$\delta^{18}\text{O}\text{‰ VSMOW}$	$\delta^{18}\text{O}\text{‰ PDB}$
CHO-TL 1	" <i>Cormohipparion</i> " sp.	KGTE 163	left upper M1	-3.7	32.7	1.8
CHO-TL 6	" <i>Cormohipparion</i> " sp.	KGTE 162	left lower M1	-4.8	31.9	0.9
CHO-TL 9	" <i>Cormohipparion</i> " sp.	KGTE 164	premolar/molar fragment	-5.3	33.3	2.3
CHOR 1 (ave)	" <i>Cormohipparion</i> " sp.		left lower P4	-6.3		
CHOR 2	" <i>Cormohipparion</i> " sp.		left upper M1	-4.6		
CHOR 3 (ave)	" <i>Cormohipparion</i> " sp.		left lower M1	-5.8		
CHOR 5	" <i>Cormohipparion</i> " sp.		left lower M2	-5.1		
CHO-BT 18	Hippopotamidae	KGTE 166	molar fragment	-12.0	23.8	-6.9
CHO-BT 30	Hippopotamidae	KGTE 165	molar fragment	-12.5	24.2	-6.5

Values are in parts per thousand difference from  $^{13}\text{C}/^{12}\text{C}$  and  $^{18}\text{O}/^{16}\text{O}$  ratios of international standards. PDB fossil carbonate for carbon; PDB and Vienna Standard Mean Ocean Water for oxygen. CHOR 1 to CHOR 5 data are from ref. 4; (ave) indicates the average of three samples of the same tooth. The Chorora type locality equid  $\delta^{13}\text{C}$  results are discussed in ref. 22 and indicate a mixed feeding diet of ~35–60%  $\text{C}_4$ . The Beticha hippopotamid  $\delta^{13}\text{C}$  values indicate that these individuals ate virtually no  $\text{C}_4$  plants. These values are lower than in the chronologically earlier (9.5–9.9 Myr old) hippopotamids (*Kenyaipotamus corydoni*) from Nakali (–9.7‰) and Samburu Hills (–6.0‰ or –5.6‰) (ref. 26), and much lower than in the later (6.5–7.4 Myr old) Hippopotaminae from the Lothagam Lower Nawata Member (–4.3‰) (ref. 26) and Toros Ménalla in Chad (–3.5‰) (ref. 49).



# A thalamic input to the nucleus accumbens mediates opiate dependence

Yingjie Zhu<sup>1</sup>, Carl F. R. Wienecke<sup>1</sup>, Gregory Nachtrab<sup>1</sup> & Xiaoke Chen<sup>1</sup>

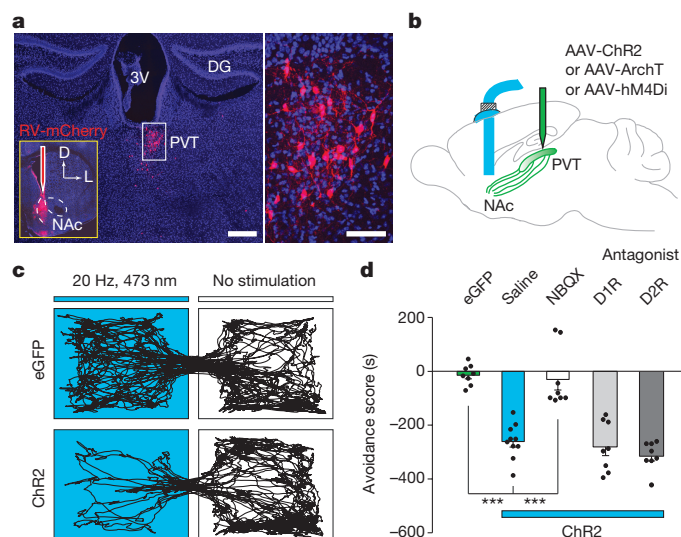
**Chronic opiate use induces opiate dependence, which is characterized by extremely unpleasant physical and emotional feelings after drug use is terminated. Both the rewarding effects of a drug and the desire to avoid withdrawal symptoms motivate continued drug use<sup>1–3</sup>, and the nucleus accumbens is important for orchestrating both processes<sup>4,5</sup>. While multiple inputs to the nucleus accumbens regulate reward<sup>6–9</sup>, little is known about the nucleus accumbens circuitry underlying withdrawal. Here we identify the paraventricular nucleus of the thalamus as a prominent input to the nucleus accumbens mediating the expression of opiate-withdrawal-induced physical signs and aversive memory. Activity in the paraventricular nucleus of the thalamus to nucleus accumbens pathway is necessary and sufficient to mediate behavioural aversion. Selectively silencing this pathway abolishes aversive symptoms in two different mouse models of opiate withdrawal. Chronic morphine exposure selectively potentiates excitatory transmission between the paraventricular nucleus of the thalamus and D2-receptor-expressing medium spiny neurons via synaptic insertion of GluA2-lacking AMPA receptors. Notably, *in vivo* optogenetic depotentiation restores normal transmission at these synapses and robustly suppresses morphine withdrawal symptoms. This links morphine-evoked pathway- and cell-type-specific plasticity in the paraventricular nucleus of the thalamus to nucleus accumbens circuit to opiate dependence, and suggests that reprogramming this circuit holds promise for treating opiate addiction.**

To systematically map brain regions that directly innervate the nucleus accumbens (NAc), we stereotactically injected a rabies virus in which the viral glycoprotein was replaced by red fluorescent protein mCherry (RV-mCherry) into the medial shell of the NAc<sup>10,11</sup>. Besides well-characterized inputs to the NAc, such as the prefrontal cortex, ventral hippocampus and basolateral amygdala (BLA)<sup>8,10,12</sup> (Extended Data Fig. 1a), we also detected mCherry-expressing neurons in the paraventricular nucleus of the thalamus (PVT) (Fig. 1a and Extended Data Fig. 1b). This result was particularly interesting because although previous studies have suggested a potentially important role for the PVT in drug-seeking behaviour, its underlying circuitry mechanism remains unknown<sup>13–15</sup>.

To characterize the PVT to NAc connection, we injected channel-rodop2-expressing adeno-associated virus (AAV-ChR2) into the PVT, then prepared acute NAc slices containing ChR2-expressing terminals from the PVT<sup>16</sup> (Extended Data Fig. 2a). Brief light stimulation (3–5 ms) elicited reliable firing of action potentials up to 20 Hz in ChR2-expressing PVT neurons (Extended Data Fig. 2b). The same stimulation also evoked robust  $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor (AMPA)-mediated excitatory postsynaptic currents (EPSCs) in medium spiny neurons (MSNs), as it was blocked by bath application of a competitive AMPAR antagonist CNQX (Extended Data Fig. 2c). Light stimulation also evoked picrotoxin-sensitive inhibitory postsynaptic currents (IPSCs) (Extended Data Fig. 3b). Because the PVT contains few, if any, GABAergic neurons (Extended Data Fig. 3a), it is likely that these IPSCs were caused by

feed-forward inhibition in a local NAc circuit. Consistent with this prediction, PVT activation-evoked IPSCs had a longer delay than that of the EPSCs, and the IPSCs were blocked by CNQX (Extended Data Fig. 3c, d).

Activation of inputs from the prefrontal cortex, ventral hippocampus and BLA to the NAc is rewarding and drives self-stimulation behaviour<sup>8,9</sup>. To directly assess the behavioural consequences of PVT→NAc circuit activity, we optogenetically activated this pathway in freely moving mice and examined their motivational valence using a real-time place preference assay (Fig. 1b, Methods and Extended Data Fig. 4). Strikingly, optogenetic activation of the PVT→NAc pathway reduced the time spent in the chamber paired with light stimulation (Fig. 1c). This indicates that, unlike other major inputs to the NAc, activation of the PVT→NAc pathway is aversive rather than rewarding.



**Figure 1 | *In vivo* optical activation of the PVT→NAc pathway evokes behavioural aversion.** **a**, Left: cluster of retrogradely labelled cells was observed in the PVT 5 days after injection of RV-mCherry into the medial shell of the NAc ( $n = 7$ ). Scale bar, 500  $\mu\text{m}$ ; inset shows the RV-mCherry injection site. Right: magnified image shows the morphology of labelled neurons in the boxed area. Scale bar, 50  $\mu\text{m}$ . D, dorsal; L, lateral; 3V, third ventricle; DG, dentate gyrus. **b**, Schematics of *in vivo* manipulation of the PVT→NAc circuit in behaving animals. **c**, Representative real-time place preference tracks illustrate light-evoked behavioural aversion in ChR2-expressing mice (bottom,  $n = 10$ ) but not in eGFP-expressing control mice (top,  $n = 8$ ). **d**, Quantification of light-evoked aversion and its effect by intra-NAc pharmacological manipulations. Intra-NAc infusions of NBQX (AMPA antagonist, 1.0  $\mu\text{g}$  in 200 nl,  $n = 8$ ) but not saline ( $n = 10$ ), SCH23390 (D1R antagonist, 0.2  $\mu\text{g}$  in 200 nl,  $n = 8$ ) or raclopride (D2R antagonist, 0.3  $\mu\text{g}$  in 200 nl,  $n = 8$ ) abolished behavioural aversion evoked by optical stimulation of the PVT→NAc fibres. One-way analysis of variance (ANOVA) ( $F_{(4,37)} = 29.61$ ,  $P < 0.0001$ ) followed by post-hoc Tukey's test. \*\*\* $P < 0.001$ . Mean  $\pm$  s.e.m.

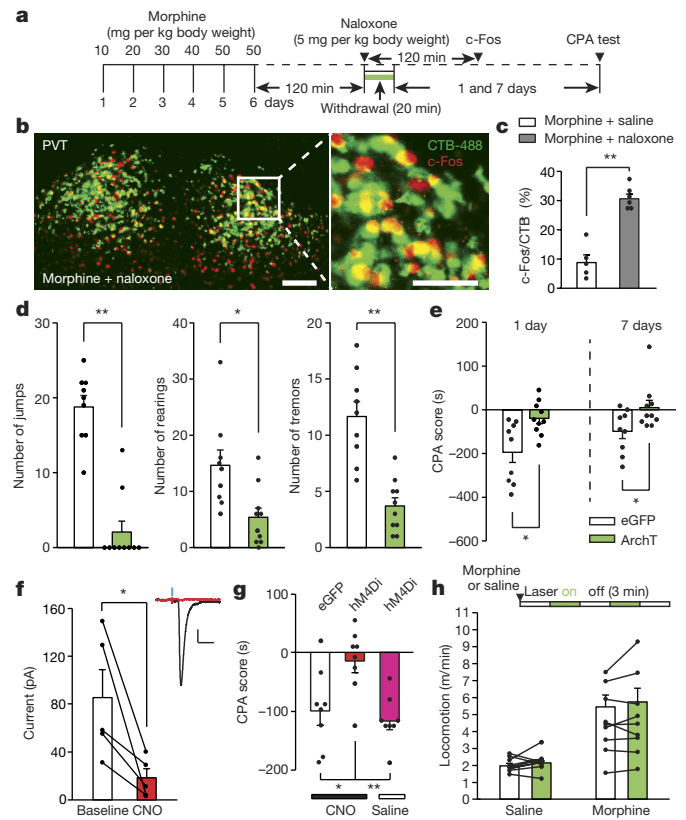
<sup>1</sup>Department of Biology, Stanford University, Stanford, California 94305, USA.

Avoidance of the light-paired chamber was dependent on local glutamatergic but not dopaminergic transmission in the NAc. Intra-NAc infusion of AMPAR antagonist NBQX abolished the avoidance behaviour, while intra-NAc infusion of D1 receptor (D1R) antagonist SCH23390 or D2 receptor (D2R) antagonist raclopride had no effect (Fig. 1d). These results demonstrate that the PVT→NAc pathway transmits negative valence, and reveal an input-specific mechanism driving motivated behaviour in the NAc.

Since activation of the PVT→NAc pathway evoked an aversive response, this pathway could be instrumental for the negative symptoms of drug withdrawal. To test this, we selectively silenced the pathway in two different models of opiate withdrawal and examined withdrawal-induced physical signs and place aversion. We used optogenetic terminal silencing by injecting an archaerhodopsin-3 (ArchT)-expressing AAV into the PVT and bilaterally implanting fibre guide cannulae in the NAc<sup>17</sup> (Extended Data Fig. 4). Two weeks after surgery, mice were rendered opiate dependent via daily intraperitoneal (i.p.) injections of morphine in their home cage with doses escalating from 10 to 50 mg per kg body weight<sup>18</sup> (Fig. 2a). Two hours after the final morphine treatment, we injected naloxone, a  $\mu$ -opioid receptor antagonist (5 mg per kg body weight, i.p.), and confined the mice to one side of a two-compartment conditioned place aversion (CPA) training chamber<sup>3</sup>. This naloxone dose evoked strong negative somatic signs and robust avoidance to the withdrawal chamber in morphine-dependent mice, but not in drug-naïve mice (Extended Data Fig. 5). Moreover, injection of naloxone in chronic morphine-treated mice elicited robust expression of c-Fos, a marker for recent neuronal activity, in the PVT<sup>NAc</sup> projection neurons, which were labelled by injecting of retrograde tracer CTB-488 into the NAc (Fig. 2b, c and Extended Data Fig. 6a, b). Remarkably, constant bilateral optogenetic silencing of the PVT→NAc pathway during naloxone-precipitated withdrawal suppressed somatic signs of opiate dependence and learned place-aversion (measured 1 and 7 days after withdrawal) (Fig. 2d, e).

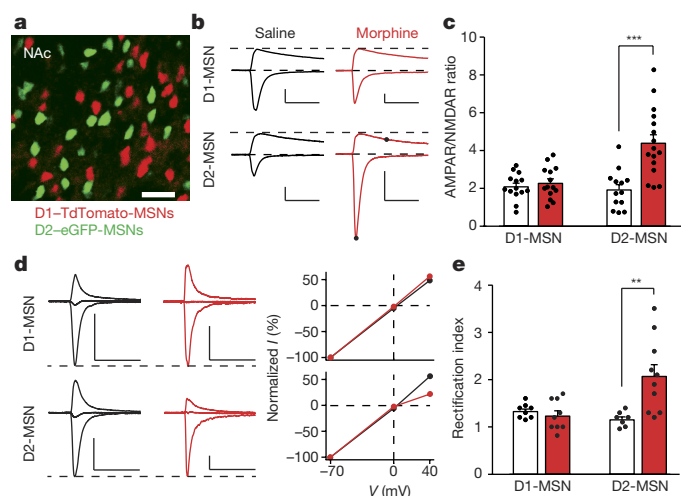
Opiate withdrawal in humans often results from cessation or reduction of opiate use, rather than blockade of opioid receptors, and this process can also be modelled in mice<sup>2,19</sup>. Spontaneous opiate withdrawal also evoked expression of c-Fos in PVT<sup>NAc</sup> projection neurons (Extended Data Fig. 6b, c). To measure aversive motivational states during spontaneous opiate withdrawal, mice were confined for 45 min to one side of a CPA training chamber 16 h after each morphine injection. After four training sessions, mice developed aversion to the withdrawal chamber. We employed inhibitory designer receptors exclusively activated by designer drugs (DREADD; AAV-hM4Di) to selectively silence the PVT→NAc pathway during each 45 min CPA training trial<sup>20,21</sup> (Fig. 2f and Methods). Local infusion of clozapine-N-oxide (CNO, 3  $\mu$ M) into the NAc before each CPA training session reduced the CPA score in AAV-hM4Di but not enhanced green fluorescent protein (eGFP)-expressing AAV (AAV-eGFP) transduced mice (Fig. 2g). PVT→NAc pathway silencing had no effect on morphine-induced locomotor activity, a measure of acute psychoactive drug effects<sup>7</sup> (Fig. 2h). Interestingly, the PVT→NAc circuit has roles beyond controlling the expression of physical and emotional negative states after opiate withdrawal. Mild footshock and i.p. injection of LiCl also evoked robust expression of c-Fos in PVT<sup>NAc</sup> projection neurons, and silencing the pathway during conditioning suppressed the ability of either stimulus to evoke behaviour aversion (Extended Data Fig. 7).

Chronic opiate use causes profound neuroadaptive changes in the NAc<sup>22,23</sup>. The NAc comprises two major subtypes of MSN, defined by the expression of either D1 or D2 dopamine receptors (D1- or D2-MSNs). D1- and D2-MSNs are proposed to play opposing roles in mediating behavioural motivation and reward learning<sup>24,25</sup>, and synaptic plasticity of these MSNs appears to be causally involved in behavioural adaptations to drug addiction and chronic pain states<sup>6,7,26</sup>. Since the PVT→NAc pathway transmits negative valence and mediates opiate withdrawal symptoms, we anticipated that chronic morphine exposure might cause plasticity of the PVT input selectively onto D2-MSNs.



**Figure 2 | The PVT→NAc pathway is required for expression of aversive withdrawal symptoms.** **a**, Experimental timeline for **b–e**. **b**, Confocal images showing that naloxone-precipitated withdrawal induced robust expression of c-Fos (red) in the PVT<sup>NAc</sup> projection neurons (green) that were retrogradely labelled by injection of CTB-488 into the medial shell of the NAc ( $n = 6$ ). Left: scale bar, 100  $\mu$ m. Right: magnified image shows the boxed area; scale bar, 50  $\mu$ m. **c**, Percentage of PVT<sup>NAc</sup> projection neurons expressing c-Fos. Naloxone (grey bar,  $n = 6$ ) but not saline (white bar,  $n = 5$ ) injection evoked significant expression of c-Fos in the PVT<sup>NAc</sup> projection neurons. Mann–Whitney  $U$ -test.  $^{**}P < 0.01$ . **d, e**, Quantification of naloxone-precipitated withdrawal behaviours and CPA score. Light stimulation in ArchT- (green bar,  $n = 10$ ) but not eGFP- (white bar,  $n = 9$ ) expressing mice during withdrawal significantly reduces the number of jump, rearing and tremor events (**d**) as well as the expression of CPA (**e**). Mann–Whitney  $U$ -test.  $^{*}P < 0.05$ ,  $^{**}P < 0.01$ . **f**, CNO (3  $\mu$ M) inhibits light-evoked synaptic current recorded from postsynaptic MSNs ( $n = 5$ ). Inset shows example light-evoked EPSC traces before (black) and after (red) perfusion of CNO. Wilcoxon signed-rank test.  $^{*}P < 0.05$ . Scale bar, 20 pA, 25 ms. **g**, Spontaneous opiate withdrawal induced CPA was reduced by local infusion of CNO in hM4Di- (red,  $n = 8$ ) but not eGFP- (black,  $n = 8$ ) expressing mice, or local infusion of saline in hM4Di- (magenta,  $n = 8$ ) expressing mice. One-way ANOVA ( $F_{(2,21)} = 7.4$ ,  $P < 0.01$ ) followed by post-hoc Tukey's test.  $^{*}P < 0.05$ ,  $^{**}P < 0.01$ . **h**, Light stimulation has no effect on locomotor velocity in either saline- ( $n = 9$ ,  $P = 0.57$ ) or morphine- ( $n = 9$ ,  $P = 0.5$ ) injected animals. Wilcoxon signed-rank test. Mean  $\pm$  s.e.m.

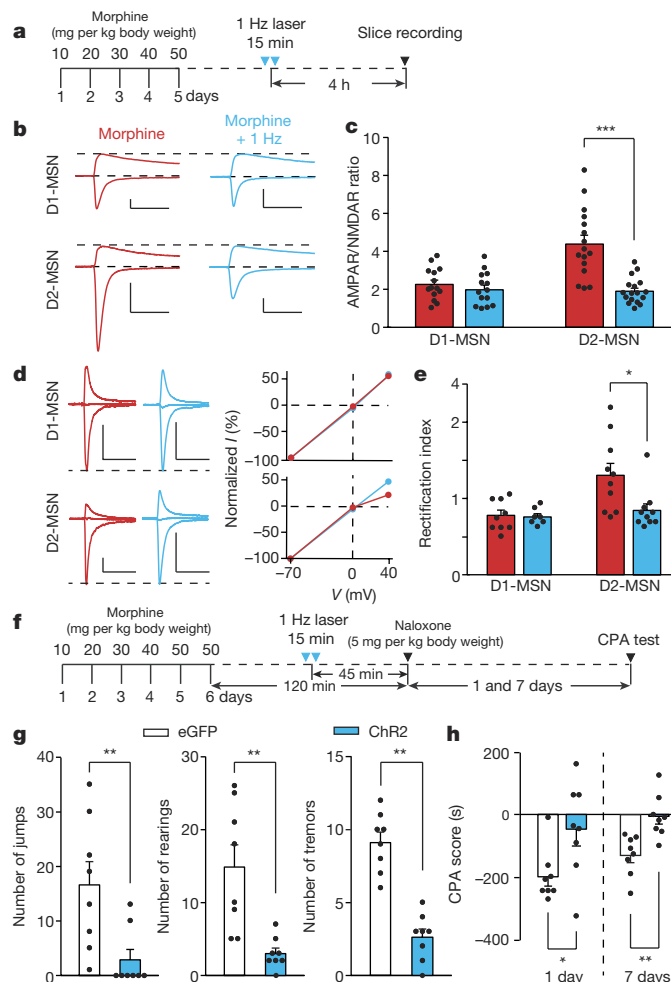
To directly examine morphine-induced synaptic plasticity of the PVT input onto identified NAc MSNs, we prepared brain slices from transgenic animals expressing fluorescent proteins under the control of the D1R or D2R promoter, with PVT neurons infected with AAV-ChR2 (Fig. 3a). To reduce experimental variability between slices, we measured the ratio of light-evoked AMPAR-mediated EPSCs to N-methyl-D-aspartate receptor (NMDAR)-mediated EPSCs (AMPAR/NMDAR ratio) in the same MSN. Consistent with our hypothesis, the escalating morphine regimen increased this ratio in D2-MSNs but not in D1-MSNs (Fig. 3b, c and Extended Data Fig. 8a, b). In contrast, presynaptic release probability was not changed in either type of MSN, as measured by the paired-pulse ratio of light-evoked EPSCs



**Figure 3 | Morphine-induced potentiation at the PVT→D2-MSN synapses.** **a**, Image of a NAc slice from a D1-TdTomato and D2-eGFP double transgenic mouse ( $n = 5$ ). Scale bar, 50  $\mu\text{m}$ . **b**, **c**, Example traces (**b**) and quantification (**c**) of light-evoked EPSCs at  $-70$  mV and  $+40$  mV show that chronic morphine treatment significantly increased the AMPAR/NMDAR ratio in D2-MSNs (saline/morphine,  $n = 13/16$  cells), but not D1-MSNs (saline/morphine,  $n = 14/14$  cells). Two-way ANOVA ( $F_{(1,53)} = 12.58$ ,  $P < 0.001$ ) followed by post-hoc Tukey's test. \*\*\* $P < 0.001$ . Filled dots indicate the current amplitude used for calculating the AMPAR/NMDAR ratio. Scale bar, 300 pA, 50 ms. **d**, **e**, Example traces (**d**, left), I/V curve (**d**, right) and quantification (**e**) of light-evoked AMPAR EPSCs at  $-70$  mV,  $0$  mV and  $+40$  mV show that morphine treatment selectively increased the rectification index of AMPAR EPSCs in D2-MSNs (saline/morphine,  $n = 7/10$  cells), but not D1-MSNs (saline/morphine,  $n = 8/9$  cells). Two-way ANOVA ( $F_{(1,30)} = 9.87$ ,  $P < 0.01$ ) followed by post-hoc Tukey's test. \*\* $P < 0.01$ . For comparison, amplitudes of AMPAR EPSCs are normalized to peaks at  $-70$  mV. Scale bar, 250 pA, 50 ms. Mean  $\pm$  s.e.m.

in morphine-dependent mice (Extended Data Fig. 9). These results suggested that chronic morphine treatment caused a change in postsynaptic function. One plausible mechanism of drug-induced postsynaptic change is the insertion of GluA2-lacking calcium-permeable AMPAR (CP-AMPA) in the affected NAc synapses<sup>27</sup>. Compared with other AMPARs, CP-AMPA has a larger conductance and show inward rectification at positive membrane potentials. In morphine-treated mice, we observed a significant increase in the rectification index selectively at PVT→D2-MSN synapses, but not at PVT→D1-MSN synapses (Fig. 3d, e). Interestingly, the AMPAR/NMDAR ratio at BLA→NAc synapses, which are part of a pathway that drives reward-seeking behaviour, was not changed by the same morphine treatment (Extended Data Fig. 8c–e). Collectively, these results indicate that the morphine regimen strengthened the PVT input selectively onto D2-MSNs via synaptic insertion of CP-AMPA.

If morphine-induced potentiation of the PVT→D2-MSN synapses is essential for the expression of opiate withdrawal symptoms, then depotentiating these synapses *in vivo* to restore its normal transmission should relieve those symptoms. To test this prediction, we employed long-term depression (LTD)-based *in vivo* manipulation of the PVT→D2-MSN synapses by photostimulating ChR2-expressing PVT terminals in the NAc at 1 Hz for 15 min<sup>6,7,28</sup> (Fig. 4a). We applied this *in vivo* optogenetic LTD induction protocol in chronic morphine-treated mice, then prepared NAc slices to record photocurrents on visually identified D1- and D2-MSNs. Indeed, this light treatment reduced the AMPAR/NMDAR ratio, and rectified the AMPAR current at PVT→D2-MSN synapses to a level comparable to baseline transmission, but did not change the paired-pulse ratio (Fig. 4b–e and Extended Data Figs 8 f, g and 9). Treating morphine-dependent mice with this optogenetic LTD protocol 45 min before naloxone injection



**Figure 4 | *In vivo* optogenetic LTD induction restores normal transmission at PVT→D2-MSN synapses and suppresses withdrawal symptoms.** **a**, Experimental timeline for **b**–**e**. **b**–**e**, *In vivo* 1 Hz photostimulation successfully normalized morphine-induced changes of AMPAR/NMDAR ratio and rectification index in D2-MSNs, but had little effect on D1-MSNs. **b**, **c**, AMPAR/NMDAR ratio of D2-MSNs (morphine/morphine + 1 Hz,  $n = 16/17$  cells) and D1-MSNs (morphine/morphine + 1 Hz,  $n = 14/14$ ). Two-way ANOVA ( $F_{(1,57)} = 13.86$ ,  $P < 0.001$ ) followed by post-hoc Tukey's test. \*\*\* $P < 0.001$ . **d**, **e**, Rectification index for AMPAR EPSCs in D2-MSNs (morphine/morphine + 1 Hz,  $n = 10/10$ ) and D1-MSNs (morphine/morphine + 1 Hz,  $n = 9/7$ ). Two-way ANOVA ( $F_{(1,32)} = 4.3$ ,  $P < 0.05$ ) followed by post-hoc Tukey's test. \* $P < 0.05$ . Scale bars, 300 pA, 50 ms (**b**); 250 pA, 50 ms (**d**). **f**, Experimental timeline for **g**, **h**. **g**, **h**, Quantification of withdrawal behaviours (**g**) and CPA score (**h**). In ChR2- (blue bar,  $n = 8$ ) but not eGFP- (white bar,  $n = 8$ ) expressing mice, *in vivo* 1 Hz stimulation suppressed naloxone-precipitated jumping, rearing and tremor events (**g**), and suppressed CPA to the withdrawal chamber (**h**). Mann–Whitney  $U$ -test. \* $P < 0.05$ , \*\* $P < 0.01$ . Mean  $\pm$  s.e.m.

(Fig. 4f) reduced the immediate expression of withdrawal behavioural symptoms and the aversive memory of the withdrawal chamber (Fig. 4g, h). This 1 Hz optogenetic stimulation had no effect on the plasticity at PVT→D1-MSN synapses (Fig. 4b–e). Together, these results establish a causal link between plasticity at PVT→D2-MSN synapses and the negative somatic and motivational states that accompany opiate withdrawal.

Complementary to previous studies that highlight the contribution of the prefrontal cortex, BLA, and ventral hippocampus inputs to the NAc in mediating drug reward and their plasticity onto D1-MSNs after chronic drug exposure, here we show that the PVT input transmits negative valence and its plasticity at PVT→D2-MSN synapses is necessary for the expression of aversive states associated with opiate withdrawal.



We further demonstrate that optogenetic restoration of normal synaptic transmission at these synapses effectively relieves withdrawal symptoms. Our optogenetic LTD protocol may inspire the development of novel treatments for opiate addiction involving deep brain stimulation to induce plasticity at relevant synapses<sup>29</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 14 March; accepted 23 December 2015.**

**Published online 3 February 2016.**

- Koob, G. F. & Le Moal, M. Drug abuse: hedonic homeostatic dysregulation. *Science* **278**, 52–58 (1997).
- Vargas-Perez, H., Ting-A-Kee, R. A., Heinmiller, A., Sturgess, J. E. & van der Kooy, D. A test of the opponent-process theory of motivation using lesions that selectively block morphine reward. *Eur. J. Neurosci.* **25**, 3713–3718 (2007).
- Wikler, A. A theory of opioid dependence. *NIDA Res. Monogr.* **30**, 174–178 (1980).
- Harris, G. C. & Aston-Jones, G. Involvement of D2 dopamine receptors in the nucleus accumbens in the opiate withdrawal syndrome. *Nature* **371**, 155–157 (1994).
- Koob, G. F., Wall, T. L. & Bloom, F. E. Nucleus accumbens as a substrate for the aversive stimulus effects of opiate withdrawal. *Psychopharmacology* **98**, 530–534 (1989).
- Pascoli, V. *et al.* Contrasting forms of cocaine-evoked plasticity control components of relapse. *Nature* **509**, 459–464 (2014).
- Pascoli, V., Turiault, M. & Lüscher, C. Reversal of cocaine-evoked synaptic potentiation resets drug-induced adaptive behaviour. *Nature* **481**, 71–75 (2012).
- Britt, J. P. *et al.* Synaptic and behavioral profile of multiple glutamatergic inputs to the nucleus accumbens. *Neuron* **76**, 790–803 (2012).
- Stuber, G. D. *et al.* Excitatory transmission from the amygdala to nucleus accumbens facilitates reward seeking. *Nature* **475**, 377–380 (2011).
- Lim, B. K., Huang, K. W., Grueter, B. A., Rothwell, P. E. & Malenka, R. C. Anhedonia requires MC4R-mediated synaptic adaptations in nucleus accumbens. *Nature* **487**, 183–189 (2012).
- Wickersham, I. R., Finke, S., Conzelmann, K. K. & Callaway, E. M. Retrograde neuronal tracing with a deletion-mutant rabies virus. *Nature Methods* **4**, 47–49 (2007).
- Sesack, S. R. & Grace, A. A. Cortico-basal ganglia reward network: microcircuitry. *Neuropsychopharmacology* **35**, 27–47 (2010).
- Martin-Fardon, R. & Boutrel, B. Orexin/hypocretin (Orx/Hcrt) transmission and drug-seeking behavior: is the paraventricular nucleus of the thalamus (PVT) part of the drug seeking circuitry? *Front. Behav. Neurosci.* **6**, 75 (2012).
- Browning, J. R., Jansen, H. T. & Sorg, B. A. Inactivation of the paraventricular thalamus abolishes the expression of cocaine conditioned place preference in rats. *Drug Alcohol Depend.* **134**, 387–390 (2014).
- James, M. H. *et al.* Cocaine- and amphetamine-regulated transcript (CART) signaling within the paraventricular thalamus modulates cocaine-seeking behaviour. *PLoS ONE* **5**, e12980 (2010).
- Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neurosci.* **8**, 1263–1268 (2005).
- Chow, B. Y. *et al.* High-performance genetically targetable optical neural silencing by light-driven proton pumps. *Nature* **463**, 98–102 (2010).
- Vanderschuren, L. J. *et al.* Morphine-induced long-term sensitization to the locomotor effects of morphine and amphetamine depends on the temporal pattern of the pretreatment regimen. *Psychopharmacology* **131**, 115–122 (1997).
- Bechara, A., Nader, K. & van der Kooy, D. Neurobiology of withdrawal motivation: evidence for two separate aversive effects produced in morphine-naive versus morphine-dependent rats by both naloxone and spontaneous withdrawal. *Behav. Neurosci.* **109**, 91–105 (1995).
- Stachniak, T. J., Ghosh, A. & Sternson, S. M. Chemogenetic synaptic silencing of neural circuits localizes a hypothalamus–midbrain pathway for feeding behavior. *Neuron* **82**, 797–808 (2014).
- Armbruster, B. N., Li, X., Pausch, M. H., Herlitze, S. & Roth, B. L. Evolving the lock to fit the key to create a family of G protein-coupled receptors potentially activated by an inert ligand. *Proc. Natl Acad. Sci. USA* **104**, 5163–5168 (2007).
- Williams, J. T., Christie, M. J. & Manzoni, O. Cellular and synaptic adaptations mediating opioid dependence. *Physiol. Rev.* **81**, 299–343 (2001).
- Hyman, S. E., Malenka, R. C. & Nestler, E. J. Neural mechanisms of addiction: the role of reward-related learning and memory. *Annu. Rev. Neurosci.* **29**, 565–598 (2006).
- Kravitz, A. V., Tye, L. D. & Kreitzer, A. C. Distinct roles for direct and indirect pathway striatal neurons in reinforcement. *Nature Neurosci.* **15**, 816–818 (2012).
- Lobo, M. K. *et al.* Cell type-specific loss of BDNF signaling mimics optogenetic control of cocaine reward. *Science* **330**, 385–390 (2010).
- Schwartz, N. *et al.* Decreased motivation during chronic pain requires long-term depression in the nucleus accumbens. *Science* **345**, 535–542 (2014).
- Conrad, K. L. *et al.* Formation of accumbens GluR2-lacking AMPA receptors mediates incubation of cocaine craving. *Nature* **454**, 118–121 (2008).
- Lee, B. R. *et al.* Maturation of silent synapses in amygdala–accumbens projection contributes to incubation of cocaine craving. *Nature Neurosci.* **16**, 1644–1651 (2013).
- Creed, M., Pascoli, V. J. & Lüscher, C. Refining deep brain stimulation to emulate optogenetic treatment of synaptic pathology. *Science* **347**, 659–664 (2015).

**Acknowledgements** We thank R. C. Malenka, K. Shen, L. Luo, T. R. Clandinin and members of the Chen laboratory for comments on the manuscript. We thank M. Asaad, J. Charalel and X. Sun for tracing and behaviour experiments. This work was supported by grants from The Whitehall Foundation, the Ajinomoto innovation alliance program, the Terman Scholarship and start-up funding from Stanford University. G.N. is supported by a training grant from the National Institute on Drug Abuse (5T32DA035165-02).

**Author Contributions** X.K.C. conceived the study. Y.J.Z., G.N. and X.K.C. designed the experiments. Y.J.Z. and C.F.R.W. conducted experiments. Y.J.Z. and X.K.C. analysed data. G.N., Y.J.Z. and X.K.C. wrote the manuscript, and all authors participated in the revision of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to X.K.C. (xkchen@stanford.edu).

## METHODS

**Animals.** Male adult (5–12 weeks) C57BL/6, Drd1a–tdTomato, Drd2–eGFP and Drd1a–tdTomato/Drd2–eGFP double BAC transgenic mice were used for all experiments<sup>10</sup>. All procedures were in accordance with the US National Institutes of Health (NIH) guidelines for the care and use of laboratory animals, and were approved by Stanford University's Administrative Panel on Laboratory Animal Care.

**Surgery.** Stereotaxic injections were performed on 5- to 8-week-old mice under ketamine and xylazine (100 mg per kg body weight and 10 mg per kg body weight, i.p.) anaesthesia using a stereotaxic instrument (BenchMARK Digital, Lecia). A small volume of concentrated virus solution was injected into the PVT (200 nl AAV, bregma –1.4 mm; lateral 0.1 mm; ventral 3.0 mm, with a 4° angle towards the midline) or medial shell of the NAc (500 nl RV-mCherry, bregma 1.1 mm; lateral 0.8 mm; ventral 4.6 mm) with a pulled glass capillary at a slow rate (100 nl min<sup>-1</sup>) using a pressure microinjector (Micro 4 system, World Precision Instruments). The injection needle was withdrawn 5 min after the end of the injection. For mice involved in behavioural experiments, a fibre-optic cannula (plastic one) was placed at least 500 µm above the medial shell of the NAc, and cemented onto the skull using dental cement (Lang Dental Manufacturing). After surgery, a dummy was inserted to keep the guide cannula from getting clogged. Mice were allowed at least 2 weeks to recover and to express the virus before behavioural training. Rabies virus-injected mice were used 5 days after injection. The injected AAV led to the expression of ChR2, ArchT or hM4Di in PVT neurons extending from Bregma –1.82 mm to –0.94 mm (Extended Data Fig. 4), which covered the area with the highest density of PVT<sup>NAc</sup> projection neurons and the highest density of neurons expressing c-Fos in response to naloxone-precipitated withdrawal (Extended Data Fig. 1b–e). AAVs used in this study were produced by the University of Pennsylvania vector core: AAV1.hSyn.eGFP.WPRE.bGH,  $1.76 \times 10^{13}$  genomic copies per millilitre; AAV1.hSyn.hChR2(H134R)-eYFP.WPRE.bGH,  $3.35 \times 10^{13}$  genomic copies per millilitre; AAV9.CAG.ArchT.GFP.WPRE.SV40,  $7.62 \times 10^{12}$  genomic copies per millilitre; and by the University of North Carolina vector core: AAV8.hSyn.hM4Di-mCherry,  $8.3 \times 10^{12}$  genomic copies per millilitre. The G-deleted Rabies-mCherry virus ( $4.5 \times 10^9$  genomic copies per millilitre) was produced by the gene transfer, targeting and therapeutics core at the Salk Institute for Biological Studies.

**Behavioural assays.** All mice used in behavioural assays were allowed to recover from surgery of AAV injection and cannula implantation for at least 3 weeks. Place preference training was performed in a custom-made two-compartment conditioned place preference (CPP) apparatus (30 cm × 25 cm × 20 cm). All locomotor behaviours, including real-time place preference, morphine-induced locomotion, and withdrawal-, LiCl-, footshock-induced place aversion were recorded for 15–20 min at 30 frames per second with a camera controlled by custom tracking software running on MATLAB (MathWorks). CPA score was calculated by subtracting the time spent in the aversive stimulus-paired side of the chamber during baseline from the time spent in the same side of the chamber during the test.

**Real-time place preference assay<sup>30</sup>.** After connecting with optical fibre, mice infected with AAV-ChR2 or AAV-eGFP were placed in the CPP training apparatus for 15 min to assess their baseline place preference. During the test, we assigned the counterbalanced side of the chamber as the stimulation side, and placed the mice in the non-stimulated side to start the experiment. When the mouse crossed to the stimulated side of the chamber, it triggered 20 Hz laser stimulation (473 nm, 20 ms pulses, 7 mW mm<sup>-2</sup> at the NAc) until the mouse crossed back to the non-stimulated side. Avoidance score was calculated by subtracting the time spent in stimulation side during baseline (without light) from the time spent in stimulation side during the test (with light).

**Naloxone-precipitated morphine withdrawal.** Mice infected with AAV-ArchT or AAV-eGFP were allowed to freely explore both sides of a custom-made CPP training apparatus for 15 min to assess their baseline place preference. Then, these mice received a single daily injection of morphine (i.p.) for 6 consecutive days with doses escalating at 10, 20, 30, 40, 50, 50 mg per kg body weight in their home cage to develop morphine dependence. Two hours after the last morphine injection, mice received an injection of naloxone (5 mg per kg body weight, i.p.) and were confined in the counterbalanced side of the CPP chamber. During the withdrawal, a constant green laser (532 nm, 5 mW mm<sup>-2</sup> at the NAc) was delivered through the cannula onto the NAc. We turned off the laser for 1 min every 5 min to avoid tissue damage<sup>31</sup>. Withdrawal symptoms were recorded for 20 min before each mouse was returned to its home cage. Physical signs (jump, rearing and tremor) were manually quantified offline. One and 7 days after naloxone injection, withdrawal mice were re-exposed to the CPP chamber and allowed to explore both sides of the CPP chamber for 15 min.

**Spontaneous morphine withdrawal<sup>19</sup>.** Mice infected with AAV-hM4Di-mCherry or AAV-eGFP were allowed to freely explore both sides of a custom-made CPP training apparatus for 15 min to assess their baseline place preference.

Then, these mice received daily injections of morphine (20 mg per kg body weight, i.p.) in their home cage. Sixteen hours after each morphine injection, mice received a local infusion of CNO (3 µM, 200 nl) or saline (200 nl) through the cannula into the NAc. 20 min later, each mouse was confined for 45 min in the counterbalanced side of the training chamber. This entire procedure was repeated once per day for 4 days. Twenty-four hours after the final training session, mice were re-exposed to the CPP chamber and allowed to explore both sides of the chamber for 15 min.

**LiCl-induced CPA<sup>32</sup>.** Mice infected with AAV-hM4Di-mCherry or AAV-eGFP were allowed to freely explore both sides of a custom-made CPP training apparatus for 15 min to assess their baseline place preference. Then, these mice were injected with saline (i.p.) and confined in the counterbalanced side of the training chamber for 1 h before return to home cage. Six hours later, the same mice were received a local infusion of CNO (3 µM, 200 nl) into the NAc 20 min before i.p. injection of LiCl (300 mg per kg body weight) and confined in the other side of the chamber for 60 min. This entire procedure was repeated once per day for 4 days. Twenty-four hours after the final training session, mice were re-exposed to the CPP chamber and allowed to explore both sides of the chamber for 15 min.

**Footshock-induced CPA<sup>32,33</sup>.** On day 1, mice infected with AAV-ArchT or AAV-eGFP were allowed to freely explore both sides of a custom-made CPP training apparatus for 15 min to assess their baseline place preference. On the second day, after connecting with optical fibre, these mice were confined in the counterbalanced side of the chamber for 15 min and received ten unescapable mild footshocks (0.5 mA, 0.5 s with 90 s inter-stimulation interval) while constant green laser (532 nm, 5 mW mm<sup>-2</sup> at the NAc) was delivered through the cannula onto the NAc. Twenty-four hours later, mice were re-exposed to the CPP chamber and allowed to explore both sides of the CPP chamber for 15 min.

**Morphine-induced locomotion.** Mice infected with AAV-ArchT were habituated to the testing apparatus (36 cm × 36 cm × 20 cm) for two daily 30 min sessions before being subjected to the behavioural test on day 3. On the test day, mice were connected with optical fibres and received a single injection of saline or morphine (30 mg per kg body weight, i.p.). Five minutes after injection, mice were placed in the testing apparatus for 15 min with laser stimulation (532 nm, 5 mW mm<sup>-2</sup> at the NAc) alternating from off to on every 3 min. We compared the averaged locomotion velocity from two laser-on sessions with the mean locomotion velocity from three laser-off sessions.

**In vivo optogenetic depotentiation.** Two weeks after recovery from AAV-ChR2 or AAV-eGFP injection and cannula implantation surgery, mice received daily injections of morphine (i.p.) for 5 days with escalating doses of 10, 20, 30, 40, 50 mg per kg body weight in their home cage. For slice recording, mice were photostimulated for 15 min with trains of 473 nm light (1 Hz, 4 ms, 900 pulses, 7 mW mm<sup>-2</sup> at the NAc) 20 h after the final morphine injection. For naloxone-precipitated morphine withdrawal, the same light stimulation was delivered 45 min before i.p. injection of naloxone.

**Electrophysiological recording.** Procedures for preparing acute brain slices and performing whole-cell recordings with optogenetic stimulations were similar to those described previously<sup>34,35</sup>. Coronal 250–300 µm slices containing the NAc were prepared using a vibratome (VT-1000S, Leica) in an ice-cold choline-based solution containing (in mM) 110 choline chloride, 2.5 KCl, 0.5 CaCl<sub>2</sub>, 7 MgCl<sub>2</sub>, 1.3 NaH<sub>2</sub>PO<sub>4</sub>, 1.3 Na-ascorbate, 0.6 Na-pyruvate, 25 glucose and 25 NaHCO<sub>3</sub>, saturated with 95% O<sub>2</sub> and 5% CO<sub>2</sub>. Slices were incubated in 32°C oxygenated artificial cerebrospinal fluid (in mM: 125 NaCl, 2.5 KCl, 2 CaCl<sub>2</sub>, 1.3 MgCl<sub>2</sub>, 1.3 NaH<sub>2</sub>PO<sub>4</sub>, 1.3 Na-ascorbate, 0.6 Na-pyruvate, 25 glucose and 25 NaHCO<sub>3</sub>) for at least 1 h before recording. Slices were transferred to a recording chamber and superfused with 2 ml min<sup>-1</sup> artificial cerebrospinal fluid. Patch pipettes (2–5 MΩ) pulled from borosilicate glass (PG10150-4, World Precision Instruments) were filled with a Cs-based low Cl<sup>-</sup> internal solution containing (in mM) 135 CsMeSO<sub>3</sub>, 10 HEPES, 1 EGTA, 3.3 QX-314, 4 Mg-ATP, 0.3 Na-GTP, 8 Na<sub>2</sub>-phosphocreatine, 290 mOsm kg<sup>-1</sup>, adjusted to pH 7.3 with CsOH. Whole-cell voltage-clamp recording was performed at room temperature (22–25°C) with a Multiclamp 700B amplifier and a Digidata 1440A (Molecular Devices). Data were sampled at 10 kHz and analysed with pClamp10 (Molecular Devices) or MATLAB (MathWorks). TdTomato-expressing D1-MSNs or eGFP-expressing D2-MSNs in the NAc were visualized using an upright fluorescent microscope (Olympus BX51WI). A blue light-emitting diode (470 nm, Thorlabs) controlled by digital commands from the Digidata 1440A was coupled to the microscope with a dual lamp house adaptor (5-UL180, Olympus) to deliver photostimulation. To record light-evoked EPSCs, 3–5 ms, 0.5–2 mW blue light was delivered through the objective to illuminate the entire field of view. To assess the efficiency of pharmacogenetic silencing of synaptic transmission, we first co-injected AAV-hM4Di and AAV-ChR2 into the PVT, then prepared NAc slices containing the ChR2 and hM4Di co-expressed terminals from the PVT. Light-evoked EPSCs were recorded from MSNs before and after perfusion with CNO (3 µM).

All experiments, except those in Extended Data Fig. 3, were performed in the presence of picrotoxin (100  $\mu$ M). For Extended Data Fig. 3, picrotoxin or CNQX (10  $\mu$ M) was perfused onto slices, and drug effects on light-evoked EPSCs/IPSCs were measured by recording traces before and 5 min after drug perfusion. Membrane potential was held at  $-70$  mV to record AMPAR-mediated current, at  $0$  mV to record GABA<sub>A</sub> receptor-mediated IPSC or at  $+40$  mV to record NMDAR-mediated current. Amplitude of NMDAR current was quantified at 50 ms after stimulus, when the contribution of the AMPAR component was minimal. AMPAR/NMDAR ratios were calculated by dividing the amplitude of the AMPAR current (peak current, at  $-70$  mV) by the amplitude of the NMDAR current. The paired-pulse ratio was calculated by dividing the second light-evoked EPSC by the first with 100 ms intervals between the two. The AMPAR current was recorded at three holding potentials ( $-70$  mV,  $0$  mV and  $+40$  mV) in the presence of D-AP5 (50  $\mu$ M). Rectification index ( $i_r$ ) of the AMPAR was calculated on the basis of the following equation:  $i_r = (I_{-70}/70)/(I_{40}/40)$ , in which  $I_{-70}$  and  $I_{40}$  were EPSC amplitudes recorded at  $-70$  mV and  $+40$  mV, respectively.

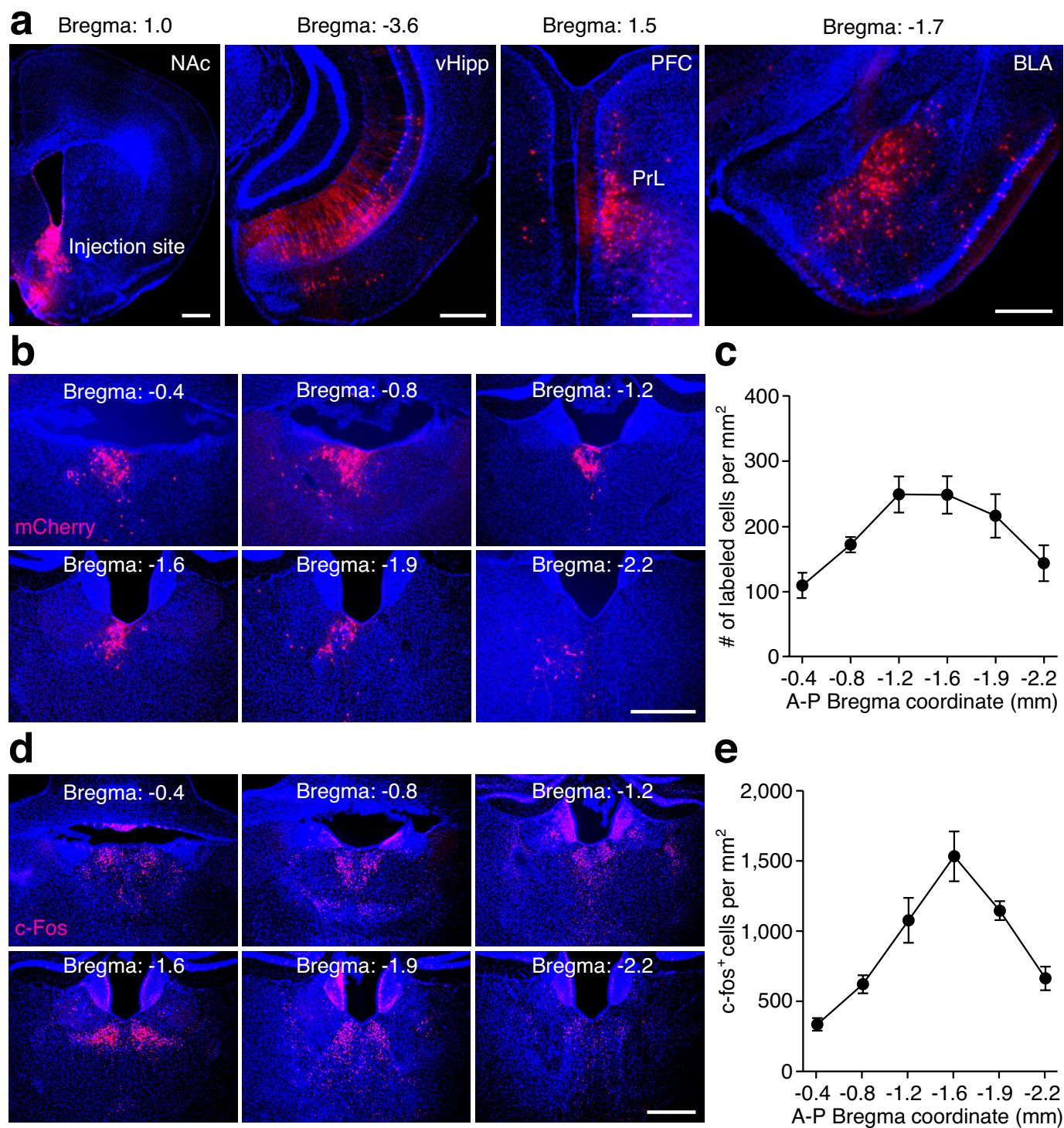
**CTB tracing and c-Fos histology.** Two to five days after bilaterally injection of CTB-488 (0.2  $\mu$ l, 0.5% in PBS) (Invitrogen, C-22841) into the medial shell of the NAc (bregma 1.1 mm; lateral 0.8 mm; ventral 4.6 mm), mice underwent naloxone-precipitated withdrawal, spontaneous withdrawal, footshock or LiCl treatment. Ninety minutes after the aversive stimuli, animals were deeply anaesthetized and perfused with 10 ml of saline (0.9%) followed by 10 ml of 4% paraformaldehyde in PBS. Coronal brain sections (50  $\mu$ M) were cut using a cryostat (Leica). Brain sections (between bregma  $-1.2$  to  $-1.9$  mm) were first washed in PBS ( $3 \times 10$  min), then blocked at room temperature with 10% NDS/0.3% Triton X-100 (PBST) and then incubated with primary anti-c-Fos antibody (Santa Cruz, SC-52G, rabbit polyclonal IgG, 1:2,000 dilution) for 3 days at  $4^\circ\text{C}$ . Brain sections were washed in PBS ( $3 \times 10$  min), followed by incubation for 2 h with fluorophore-conjugated secondary antibody (1:1,000 in 5% NDS PBST) and finally counterstained with

Hoescht (1:10,000, ThermoFisher Scientific). Images were obtained using a Zeiss 510 confocal microscope or a Zeiss epifluorescence microscope, and analysed by an individual blind to the identity of experimental groups.

**Statistics.** No statistics were used to predetermine sample size. However, our sample sizes were similar to those reported in previous publications<sup>7–9</sup>. Statistical methods are indicated when used, and statistical analyses of main figures are provided in Extended Data Table 1. All analyses were performed using MATLAB (MathWorks) or Prism. No method of randomization was used in any of the experiments. For ANOVA analyses, the variances were similar as determined by Brown–Forsythe test. Experimenters were not blind to group allocation in behavioural experiments, but CPA score and locomotion were measured automatically by custom tracking software running on MATLAB (MathWorks). All animals that finished the entire behavioural training and testing were included in analysis. Data are presented as mean  $\pm$  s.e.m.

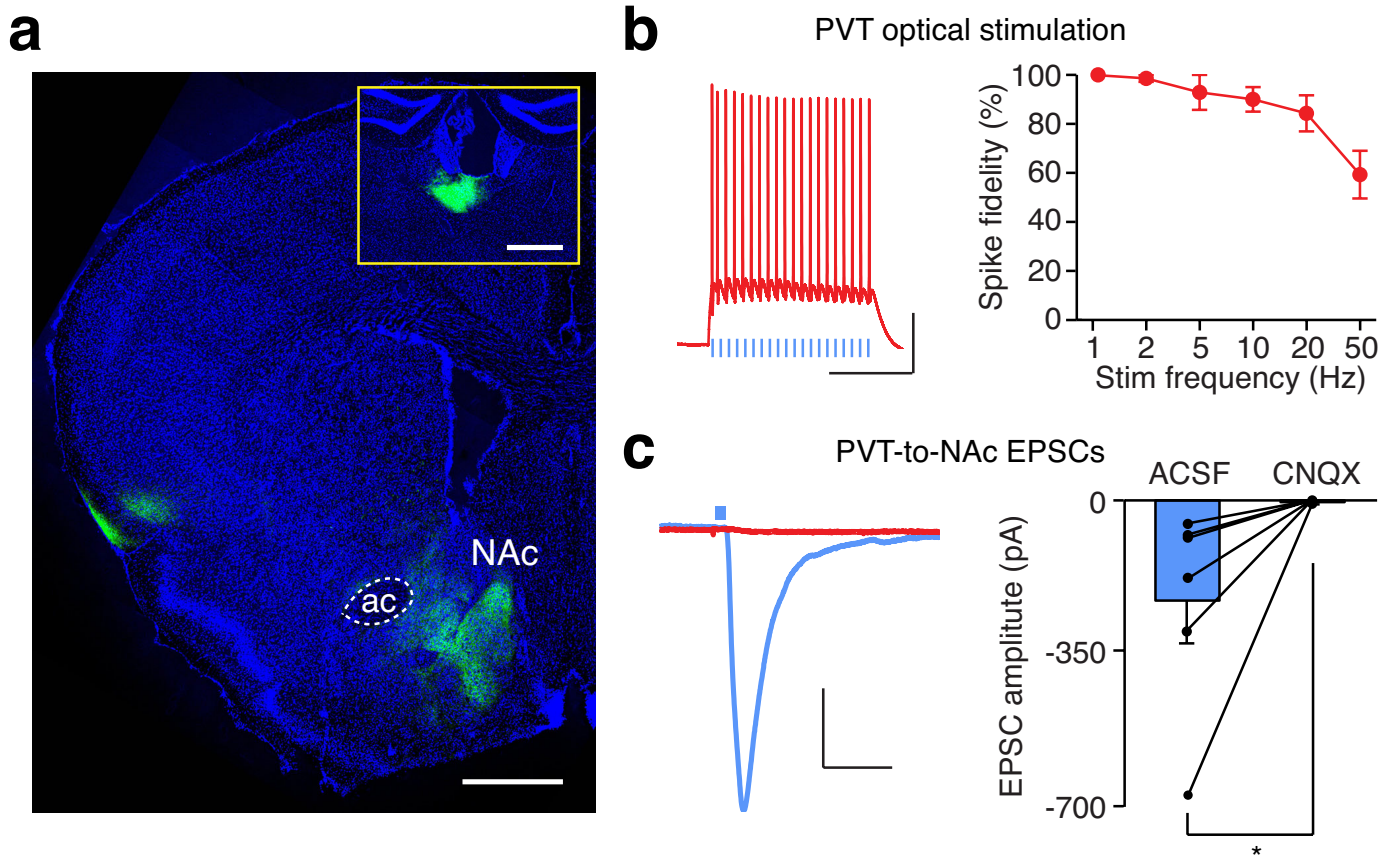
30. Stamatakis, A. M. & Stuber, G. D. Activation of lateral habenula inputs to the ventral midbrain promotes behavioral avoidance. *Nature Neurosci.* **15**, 1105–1107 (2012).
31. Tye, K. M. *et al.* Amygdala circuitry mediating reversible and bidirectional control of anxiety. *Nature* **471**, 358–362 (2011).
32. Yasoshima, Y., Scott, T. R. & Yamamoto, T. Differential activation of anterior and midline thalamic nuclei following retrieval of aversively motivated learning tasks. *Neuroscience* **146**, 922–930 (2007).
33. Wolff, S. B. *et al.* Amygdala interneuron subtypes control fear learning through disinhibition. *Nature* **509**, 453–458 (2014).
34. Ren, J. *et al.* Habenula “cholinergic” neurons co-release glutamate and acetylcholine and activate postsynaptic neurons via distinct transmission modes. *Neuron* **69**, 445–452 (2011).
35. Petreanu, L., Huber, D., Sobczyk, A. & Svoboda, K. Channelrhodopsin-2-assisted circuit mapping of long-range callosal projections. *Nature Neurosci.* **10**, 663–668 (2007).





**Extended Data Figure 1 | Retrograde tracing with rabies virus labels brain areas innervating the NAc and naloxone-precipitated opiate withdrawal induces c-Fos expression in the PVT.** **a**, Representative images show the injection site in the medial shell of the NAc, and retrograde labelling with mCherry in the ventral hippocampus (vHipp), prelimbic cortex (PrL) and basolateral amygdala (BLA). **b**, Representative images show retrograde labelling with mCherry in the PVT, ranging from Bregma -0.4 to -2.2 mm. **c**, Distribution of the density of retrogradely

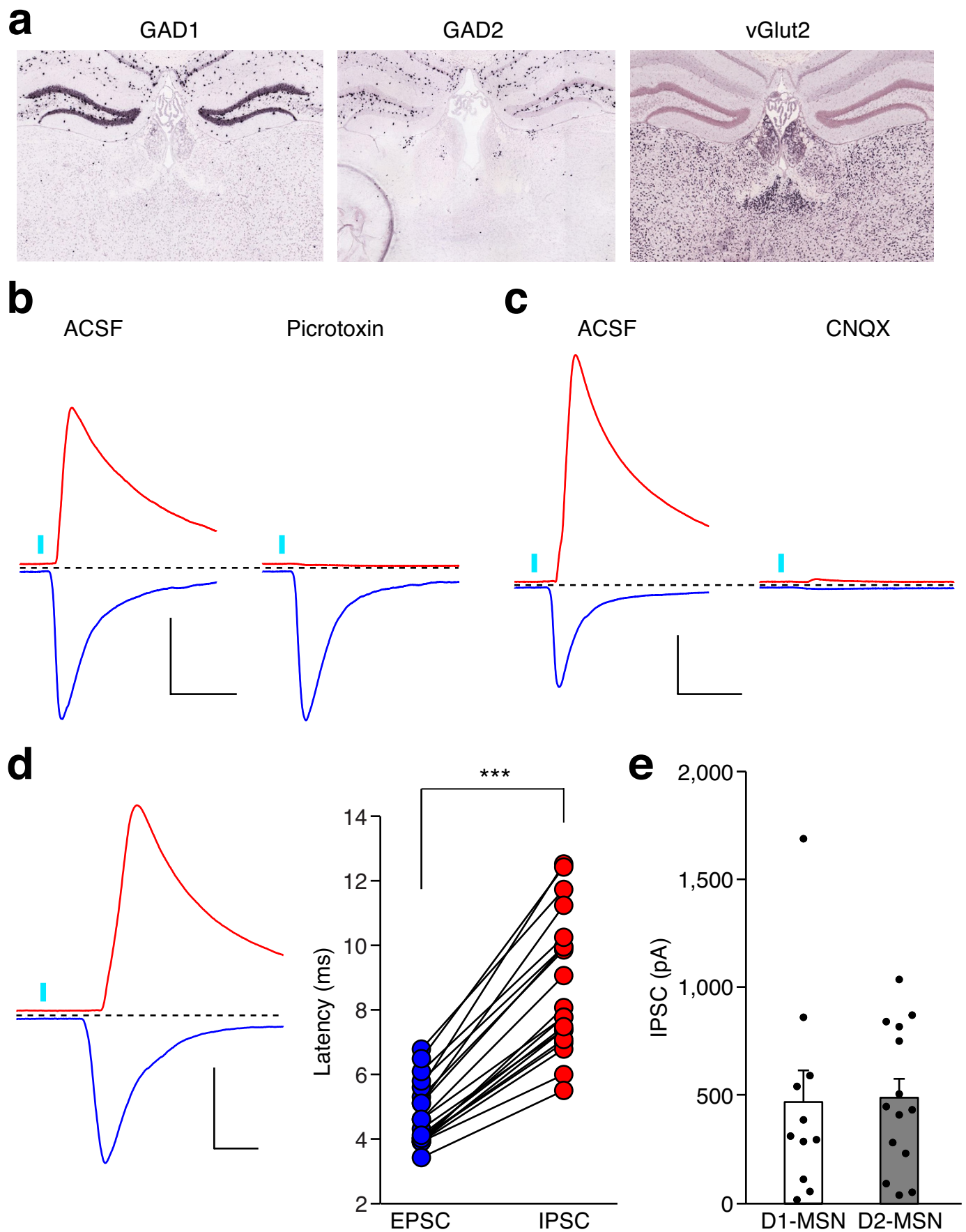
labelled neurons in the PVT ( $n = 7$ ). **d**, Representative images show c-Fos expression in the PVT induced by naloxone-precipitated opiate withdrawal. **e**, Distribution of the density of c-Fos expressing neurons in the PVT ( $n = 5$ ). Area of the PVT at different anterior-posterior coordinates are determined on the basis of the brain atlas published by Allen Institute for Brain Science (available from [http://mouse.brain-map.org/experiment/thumbnails/100142143?image\\_type=atlas](http://mouse.brain-map.org/experiment/thumbnails/100142143?image_type=atlas)). Scale bars, 500  $\mu$ m. Mean  $\pm$  s.e.m.



**Extended Data Figure 2 | Optogenetic targeting of glutamatergic transmission between the PVT and the NAc.** **a**, A confocal image shows ChR2-eYFP- (enhanced yellow fluorescent protein) expressing terminals from the PVT in the medial shell of the NAc. Inset, injection site in the PVT. Scale bars, 500  $\mu$ m. **b**, Example traces (left) and quantification (right) of action potential firing evoked by light stimulation at different

frequencies in ChR2-expressing PVT neurons ( $n = 7$ ). Scale bar, 25 mV, 250 ms. **c**, Example traces (left) and quantification (right) of photo-evoked EPSCs recorded from the NAc before and after superfusion of CNQX (10  $\mu$ M,  $n = 6$ ). Scale bar, 50 pA, 25 ms. Wilcoxon signed-rank test. \* $P < 0.05$ . Mean  $\pm$  s.e.m.





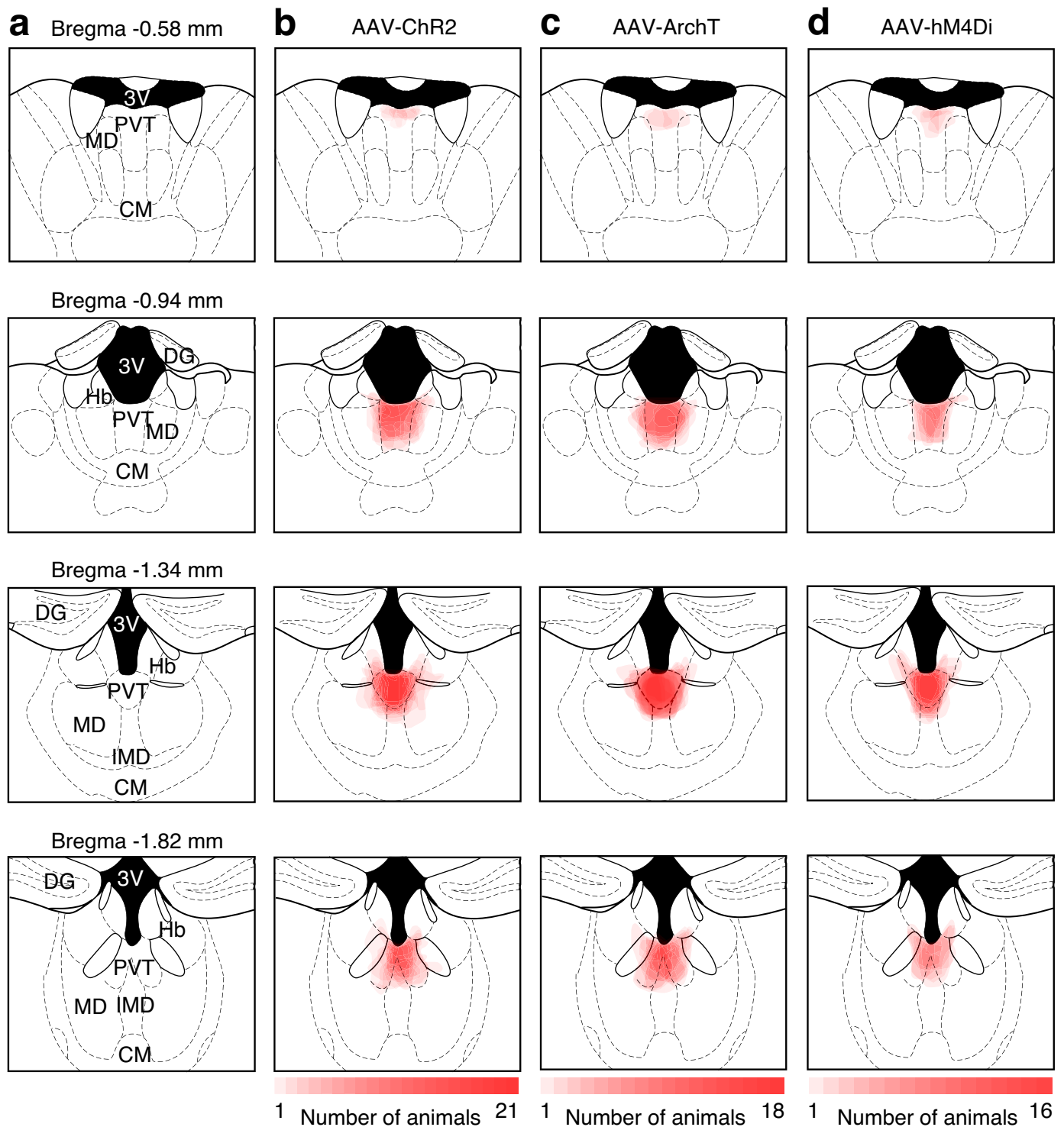
Extended Data Figure 3 | See next page for figure caption.



**Extended Data Figure 3 | Optogenetic activation of the PVT input evoked feed-forward inhibition onto MSNs in the NAc.**

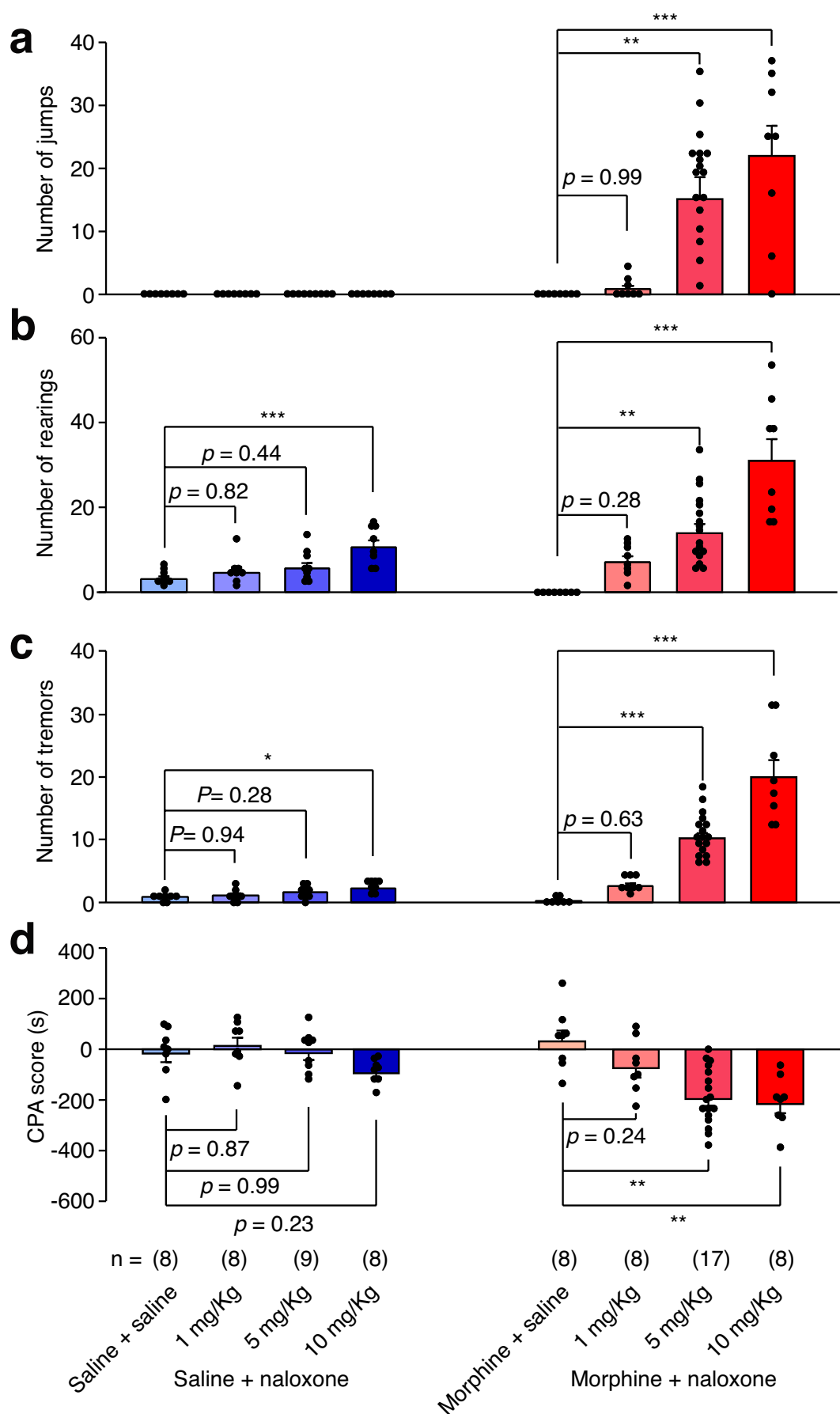
**a**, *In situ* hybridization with GAD1 (<http://mouse.brain-map.org/gene/show/14191>), GAD2 (<http://mouse.brain-map.org/experiment/show/79591669>) and vGlut2 (<http://mouse.brain-map.org/experiment/show/73818754>) in the PVT (©2014 Allen Institute for Brain Science). **b, c**, Superfusion of a GABA<sub>A</sub> receptor antagonist (picrotoxin, 100  $\mu$ M) selectively blocked IPSCs (**b**) ( $n=4$ ), whereas superfusion of an AMPAR

antagonist (CNQX, 10  $\mu$ M) blocked both IPSCs and EPSCs (**c**) ( $n=5$ ). Scale bar, 400 pA, 30 ms. **d**, Example trace (left) and quantification (right) of the onset latency for photo-evoked IPSCs and EPSCs in the same neuron ( $n=20$ ). Scale bar, 400 pA, 10 ms. Wilcoxon signed-rank test. \*\*\* $P < 0.001$ . **e**, Averaged amplitude of photo-evoked IPSCs in D1-MSNs ( $n=11$ ) and D2-MSNs ( $n=14$ ). Mann-Whitney *U*-test,  $P=0.67$ . Mean  $\pm$  s.e.m.



**Extended Data Figure 4 | Location of ChR2, ArchT and hM4Di expression in the PVT.** **a**, Schematics showing the PVT and its surrounding brain nucleus from anterior to posterior locations. 3V, third ventricle; MD, mediodorsal thalamus; CM, central middle thalamus; DG, dentate gyrus; Hb, habenula; IMD, intermediodorsal thalamus. **b**, Overlay of ChR2-YFP expression in 21 mice, including 13 mice in

Fig. 1d and 8 mice in Fig. 4g, h. **c**, Overlay of ArchT-YFP expression in 18 mice, including 10 mice in Fig. 2d and 8 mice in Extended Data Fig. 7c. **d**, Overlay of hM4Di-mCherry expression in 16 mice, including 8 mice in Fig. 2g and 8 mice in Extended Data Fig. 7d. Intensity of red colour is proportional to the number of mice expressing virus in the marked area.

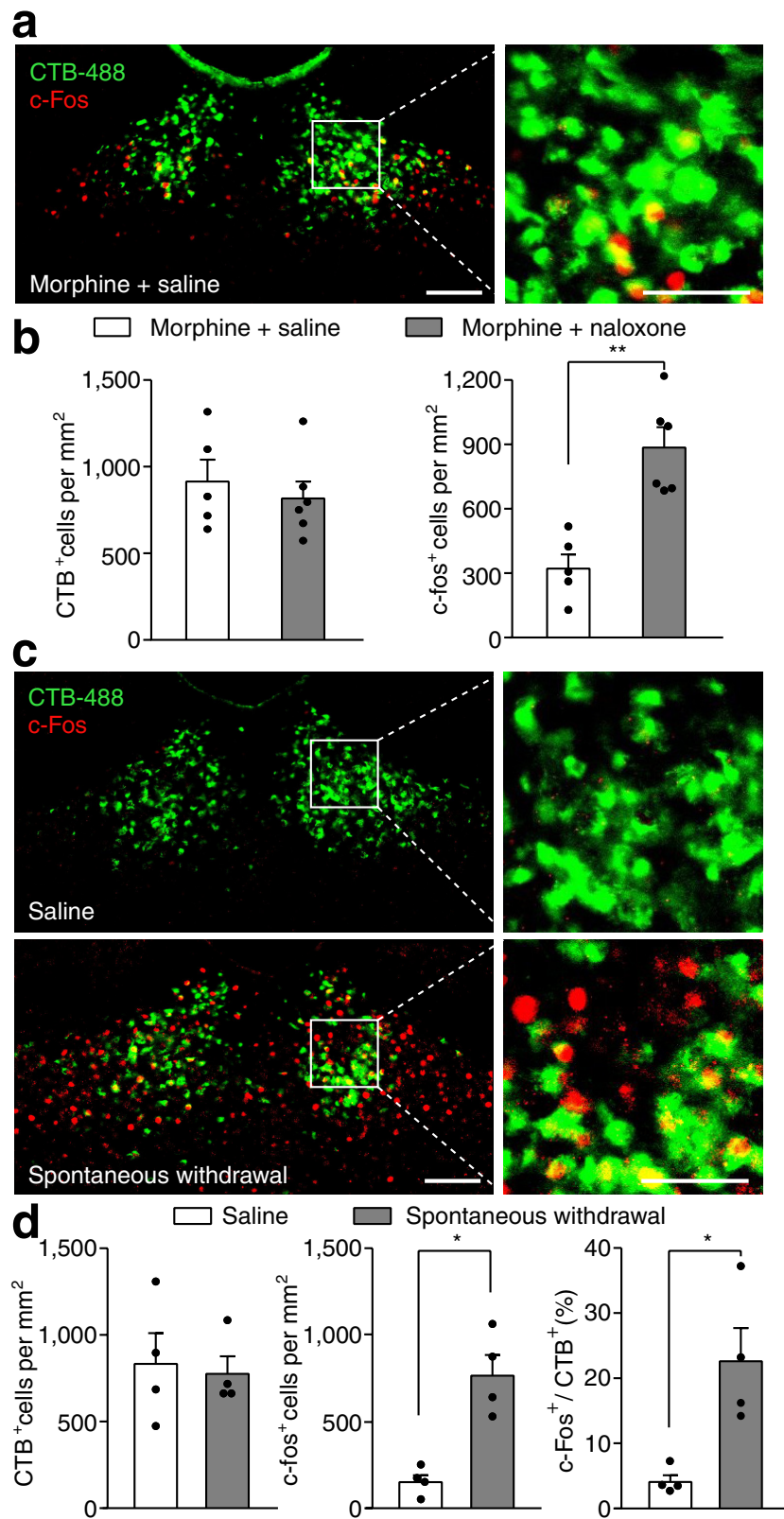


Extended Data Figure 5 | See next page for figure caption.



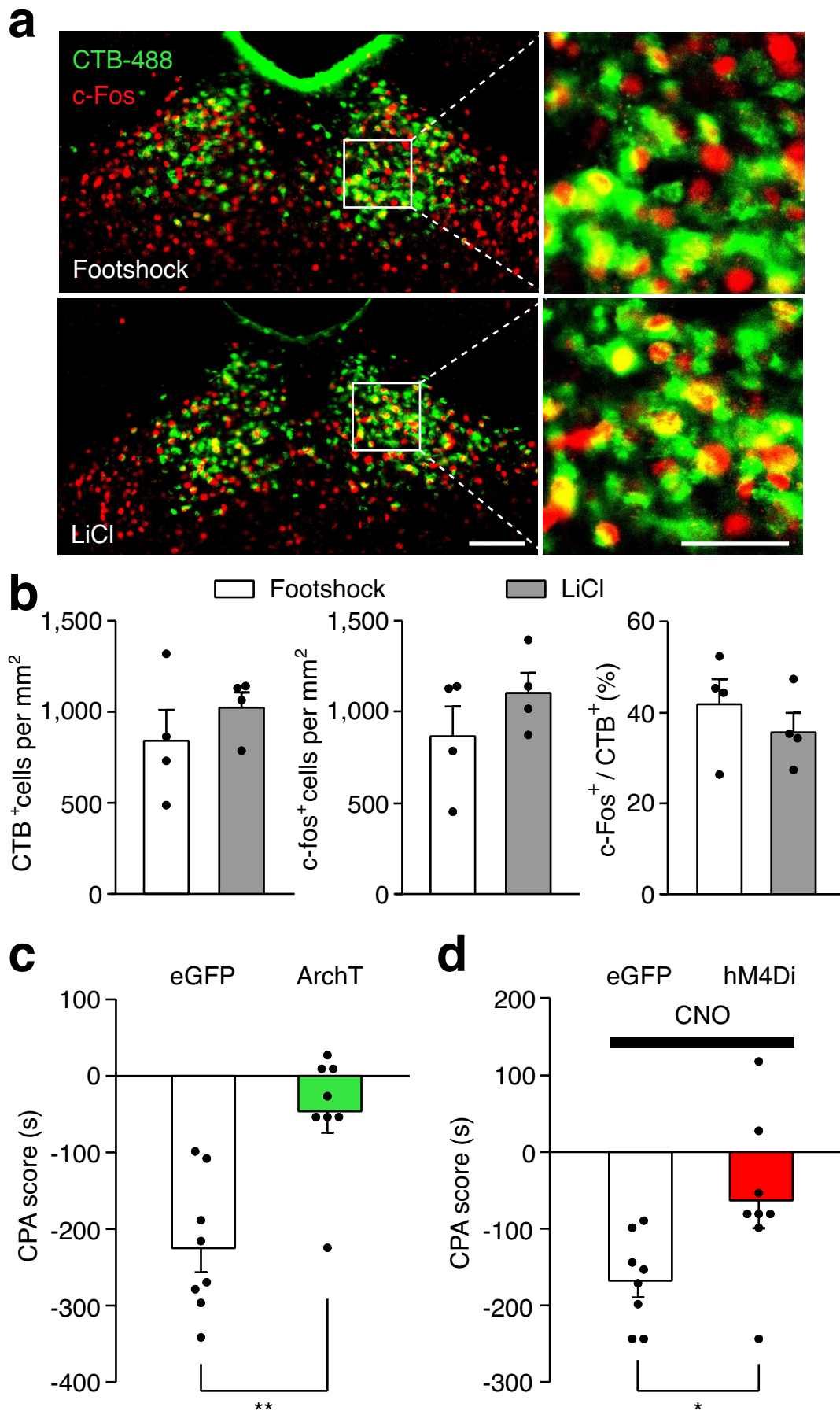
**Extended Data Figure 5 | Dose–response analysis of naloxone-precipitated withdrawal symptoms.** Concentration of naloxone was determined by  $2 \times 2$  factorial design. Different doses of naloxone (0, 1, 5, 10 mg per kg body weight) were injected in chronic saline- (blue) or morphine- (red) treated mice. Behavioural signs of withdrawal including jump (a), rearing (b) and tremor (c) were recorded for 20 min immediately after naloxone injection. CPA (d) tests were performed 24 h after withdrawal. One-way ANOVA followed by Tukey's test. Jumping:

morphine + naloxone group  $F_{(3,36)} = 9.93$ ,  $P < 0.0001$ ; rearing: saline + naloxone group  $F_{(3,36)} = 7.07$ ,  $P < 0.01$ ; morphine + naloxone group  $F_{(3,36)} = 22.98$ ,  $P < 0.0001$ ; tremor: saline + naloxone group  $F_{(3,29)} = 3.74$ ,  $P < 0.05$ ; morphine + naloxone group  $F_{(3,36)} = 40.48$ ,  $P < 0.0001$ ; CPA: saline + naloxone group  $F_{(3,29)} = 2.67$ ,  $P = 0.066$ ; morphine + naloxone group  $F_{(3,36)} = 9.93$ ,  $P < 0.0001$ ;  $P$  values for post-hoc Tukey's test are indicated on each comparison pair. Mean  $\pm$  s.e.m.



**Extended Data Figure 6 | Opiate withdrawal induced c-Fos expression in the PVT<sup>Nac</sup> projection neurons.** **a**, Example confocal image shows expression of c-Fos in a small number of PVT<sup>Nac</sup> projection neurons after injection of saline into chronic morphine treated mice. Left: scale bar, 100  $\mu$ m. Right: magnified image shows the boxed area; scale bar, 50  $\mu$ m. **b**, Quantification of CTB- (left) and c-Fos- (right) positive cells per square millimetre after injection of saline (white bar,  $n = 5$ ) or naloxone (grey bar,  $n = 6$ ) into chronic morphine-treated mice. **c**, Example confocal images

show that spontaneous withdrawal from morphine (bottom) but not saline (top) treatment increases the expression of c-Fos in PVT<sup>Nac</sup> projection neurons. Left: scale bar, 100  $\mu$ m. Right: magnified image shows the boxed area; scale bar, 50  $\mu$ m. **d**, Quantification of CTB- (left) and c-Fos- (middle) positive cells per square millimetre and percentage of PVT<sup>Nac</sup> projection neurons (right) expressing c-Fos induced by spontaneous withdrawal from morphine (grey bar,  $n = 4$ ) or saline (white bar,  $n = 4$ ). Mann-Whitney  $U$ -test. \* $P < 0.05$ , \*\* $P < 0.01$ . Mean  $\pm$  s.e.m.



Extended Data Figure 7 | See next page for figure caption.



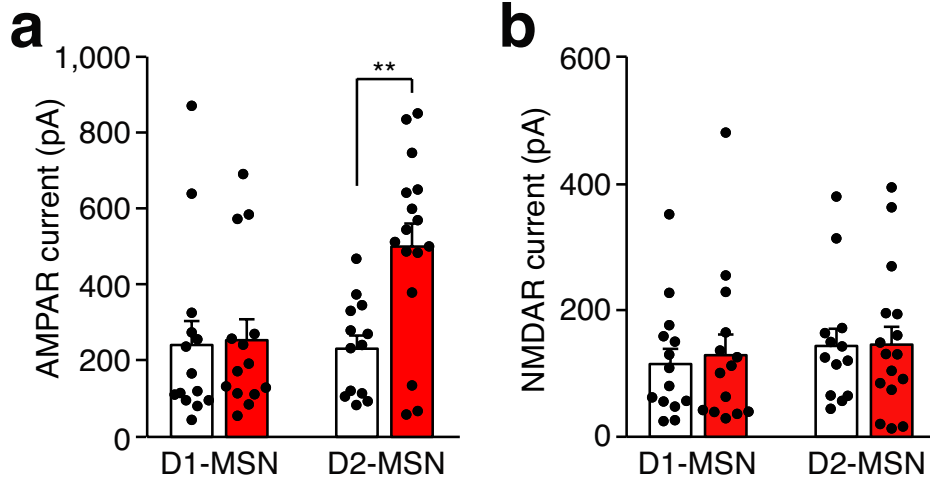
**Extended Data Figure 7 | The PVT→NAc pathway is required for expression of behavioural aversion to footshock and LiCl injection.**

**a**, Confocal images show robust expression of c-Fos (red) in the PVT<sup>NAc</sup> projection neurons (green) after footshock (top) and LiCl injection (bottom). Left: scale bar, 100  $\mu$ m. Right: magnified image shows the boxed area; scale bar, 50  $\mu$ m. **b**, Quantification of CTB- (left) and c-Fos- (middle) positive cells per square millimetre and percentage of PVT<sup>NAc</sup> projection

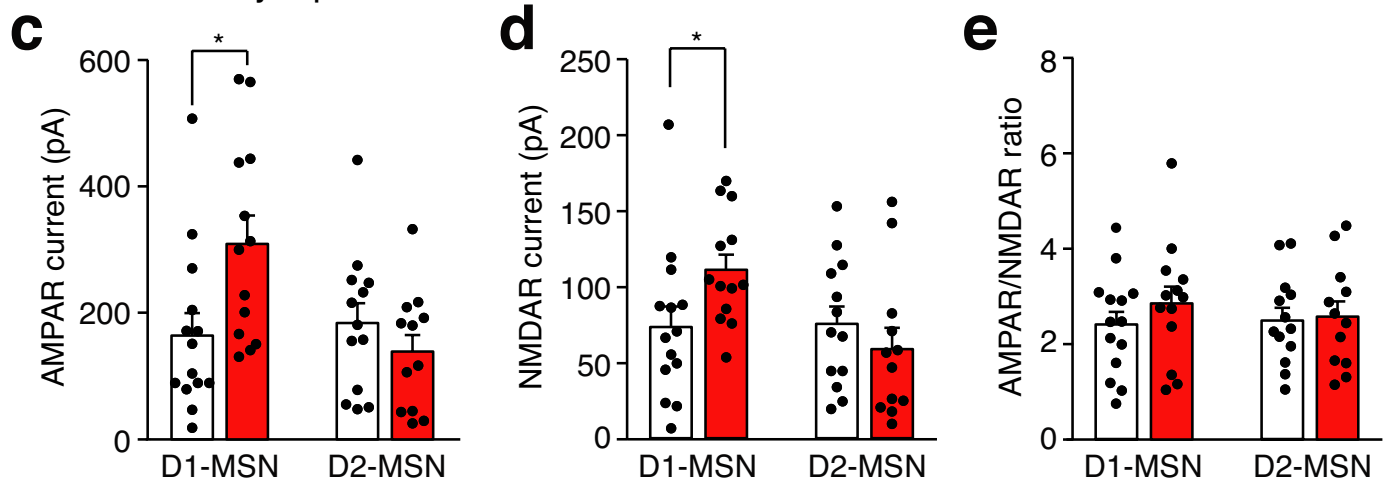
neurons (right) expressing c-Fos induced by footshock (white bar,  $n = 4$ ) or LiCl injection (grey bar,  $n = 4$ ). **c**, Light stimulation suppressed the expression of footshock-induced CPA in ArchT- ( $n = 8$ ) but not eGFP- ( $n = 8$ ) expressing mice. **d**, Local infusion of CNO reduced the expression of LiCl-induced CPA in hM4Di- ( $n = 8$ ) but not eGFP- ( $n = 8$ ) expressing mice. Mann–Whitney  $U$ -test. \* $P < 0.05$ , \*\* $P < 0.01$ . Mean  $\pm$  s.e.m.

## PVT-to-NAc synapses

□ Saline      ■ Morphine

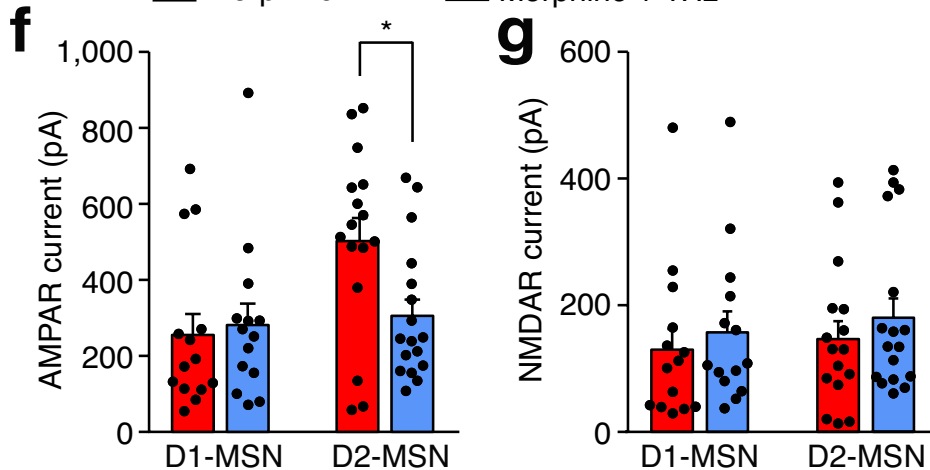


## BLA-to-NAc synapses



## Depotential of the PVT-to-NAc synapses

■ Morphine      ■ Morphine + 1Hz

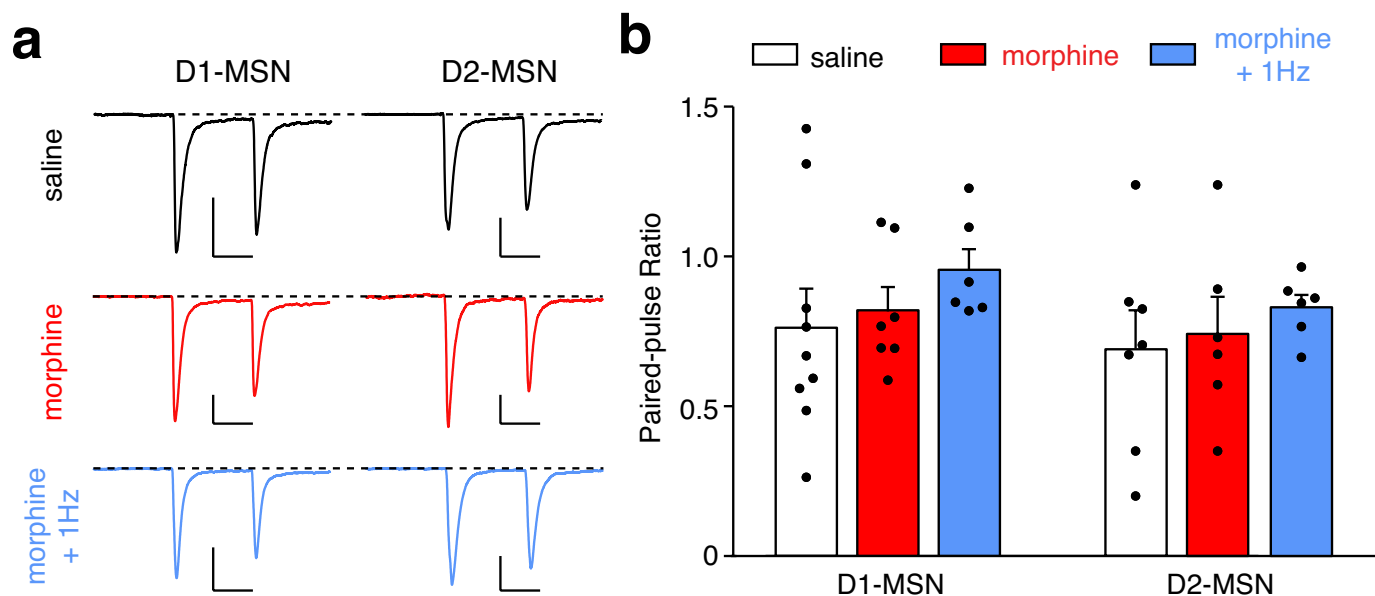


Extended Data Figure 8 | See next page for figure caption.

**Extended Data Figure 8 | Effects of chronic morphine treatment and *in vivo* optogenetic LTD induction on the strength of AMPAR and NMDAR current in MSNs receiving input from the PVT or BLA.** **a, b**, In the PVT→NAc pathway, chronic morphine treatment specifically increases AMPAR but not NMDAR current on D2-MSNs (saline/morphine,  $n = 13/16$  cells) but not D1-MSNs (saline/morphine,  $n = 14/14$  cells). Two-way ANOVA (AMPA:  $F_{(1,53)} = 5.24$ ,  $P < 0.05$ ; NMDAR:  $F_{(1,53)} = 0.04$ ,  $P = 0.83$ ) followed by Tukey's test.  $**P < 0.01$ . **c, d**, In the BLA→NAc pathway, chronic morphine treatment increases both AMPAR current (**c**) and NMDAR current (**d**) on D1-MSNs (saline/morphine,  $n = 14/13$  cells) but not D2-MSNs (saline/morphine,  $n = 13/12$

cells). Two-way ANOVA (AMPA:  $F_{(1,51)} = 7.06$ ,  $P < 0.05$ ; NMDAR:  $F_{(1,51)} = 0.35$ ,  $P = 0.55$ ) followed by Tukey's test.  $*P < 0.05$ . **e**, Chronic morphine treatment has no effect on the AMPAR/NMDAR ratio in either D1-MSNs or D2-MSNs in the BLA→NAc pathway. Two-way ANOVA ( $F_{(1,51)} = 7.06$ ,  $P < 0.05$ ). **f, g**, In the PVT→NAc pathway, *in vivo* optogenetic stimulation (4 ms, 1 Hz, 900 pulses) in morphine-treated mice specifically decreases AMPAR but not NMDAR current in D2-MSNs (morphine/morphine + 1 Hz,  $n = 16/17$  cells) but not D1-MSNs (morphine/morphine + 1 Hz,  $n = 14/14$  cells). Two-way ANOVA (AMPA:  $F_{(1,57)} = 4.24$ ,  $P = 0.04$ ; NMDAR:  $F_{(1,57)} = 0.01$ ,  $P = 0.92$ ) followed by Tukey's test.  $*P < 0.05$ . Mean  $\pm$  s.e.m.





**Extended Data Figure 9 | Chronic morphine treatment and *in vivo* optogenetic LTD induction does not affect paired-pulse ratio of MSNs receiving PVT input. a, b, Example traces (a) and quantification (b) of paired-pulse ratio of photo-evoked EPSCs in D1- and D2-MSNs. An escalating regimen of morphine treatment and *in vivo* optogenetic**

stimulation (4 ms, 1 Hz, 900 pulses) in morphine-treated mice had no obvious effect on the paired-pulse ratio of MSNs receiving PVT input (two-way ANOVA,  $F_{(2,35)} = 0.02$ ,  $P = 0.97$ ). D1-MSNs: saline/morphine/morphine + 1 Hz,  $n = 9/7/6$ ; D2-MSNs: saline/morphine/morphine + 1 Hz,  $n = 7/6/6$ . Scale bars, 200 pA, 50 ms. Mean  $\pm$  s.e.m.

Extended Data Table 1 | Statistical analysis for Figs 1–4

Figure	Groups (Ns refer to animals in behavior and c-fos experiments and cells in electrophysiology experiments)	Statistical Analysis
1d (RTPP)	eGFP control (n = 8) ChR2 + saline (n = 10) ChR2 + NBQX (n = 8) ChR2 + D1R antagonist (n = 8) ChR2 + D2R antagonist (n = 8)	One-way ANOVA: $F(4, 37) = 29.61, p < 0.0001$  Post-hoc Tukey's test: ChR2 + saline vs. GFP control, $p < 0.001$ ChR2 + saline vs. ChR2 + NBQX, $p < 0.001$ ChR2 + saline vs. ChR2 + D1R antagonist, $p = 0.98$ ChR2 + saline vs. ChR2 + D2R antagonist, $p = 0.58$
2c (c-fos)	Morphine + saline (n = 5) Morphine + naloxone (n = 6)	Mann-Whitney <i>U</i> -test Morphine + saline vs. Morphine + naloxone, $p < 0.01$
2d (Behavioral signs)	eGFP control (n = 9) ArchT silencing (n = 10)	Mann-Whitney <i>U</i> -test Jump: GFP vs. Arch, $p < 0.001$ Rearing: GFP vs. Arch, $p < 0.01$ Tremor: GFP vs. Arch, $p < 0.001$
2e (CPA)	eGFP control (n = 9) ArchT silencing (n = 10)	Mann-Whitney <i>U</i> -test 1d CPA: GFP vs. Arch, $p < 0.05$ 7d CPA: GFP vs. Arch, $p < 0.05$
2f (hM4Di silencing current)	Baseline (n = 5) CNO (n = 5)	Wilcoxon signed-rank test Baseline vs. CNO, $p < 0.05$
2g (Spontaneous withdrawal)	eGFP + CNO (n = 8) hM4D + CNO (n = 8) hM4D + saline (n = 8)	One-way ANOVA: $F(2, 21) = 7.4, p < 0.01$  Post-hoc Tukey's test: hM4D + CNO vs. GFP + CNO, $p < 0.05$ hM4D + CNO vs. hM4D + saline, $p < 0.01$ hM4D + saline vs. GFP + CNO, $p = 0.55$
2h (Locomotion)	eGFP control (n = 9) ArchT silencing (n = 9)	Wilcoxon signed-rank test Saline: GFP vs. Arch, $p = 0.57$ Morphine: GFP vs. Arch, $p = 0.5$
3c (AN ratio)	Saline D1 (n = 14) Saline D2 (n = 13) Morphine D1 (n = 14) Morphine D2 (n = 16)	Two-way factorial ANOVA Group (saline, morphine) x Cell (D1, D2) Group: $F(1,53) = 17.01, p < 0.001$ Cell: $F(1,53) = 9.18, p < 0.01$ Group X Cell: $F(1,53) = 12.58, p < 0.001$  Post-hoc Tukey's test: saline D1 vs morphine D1, $p = 0.97$ saline D2 vs morphine D2, $p < 0.001$
3e (Rectification index)	Saline D1 (n = 8) Saline D2 (n = 7) Morphine D1 (n = 9) Morphine D2 (n = 10)	Two-way factorial ANOVA Group (saline, morphine) x Cell (D1, D2) Group: $F(1,30) = 6.65, p < 0.05$ Cell: $F(1,30) = 4.32, p < 0.05$ Group X Cell: $F(1,30) = 9.87, p < 0.01$  Post-hoc Tukey's test: saline D1 vs morphine D1, $p = 0.97$ saline D2 vs morphine D2, $p < 0.01$
4c (AN ratio)	Morphine D1 (n = 14) Morphine D2 (n = 16) Morphine + 1Hz, D1 (n = 14) Morphine + 1Hz, D2 (n = 17)	Two-way factorial ANOVA Group (morphine, morphine + 1Hz) x Cell (D1, D2) Group: $F(1,57) = 21.59, p < 0.0001$ Cell: $F(1,57) = 11.57, p < 0.01$ Group X Cell: $F(1,57) = 13.86, p < 0.001$  Post-hoc Tukey's test: Morphine D1 vs morphine 1Hz D1, $p = 0.97$ Morphine D2 vs morphine 1Hz D2, $p < 0.001$
4e (Rectification index)	morphine D1 (n = 9) morphine D2 (n = 10) depotentiation D1 (n = 7) depotentiation D2 (n = 10)	Two-way factorial ANOVA Group (morphine, morphine 1Hz) x Cell (D1, D2) Group: $F(1,32) = 5.12, p < 0.05$ Cell: $F(1,32) = 8.23, p < 0.01$ Group X Cell: $F(1,32) = 4.3, p < 0.05$  Post-hoc Tukey's test: saline D1 vs morphine D1, $p = 0.92$ saline D2 vs morphine D2, $p < 0.05$
4g (Behavioral signs)	eGFP control (n = 8) ChR2 depotentiation (n = 8)	Mann-Whitney <i>U</i> -test Jump: GFP vs. ChR2, $p < 0.01$ Rearing: GFP vs. ChR2, $p < 0.01$ Tremor: GFP vs. ChR2, $p < 0.01$
4h (CPA)	eGFP control (n = 8) ChR2 depotentiation (n = 8)	Mann-Whitney <i>U</i> -test 1d CPA: GFP vs. ChR2, $p < 0.05$ 7d CPA: GFP vs. ChR2, $p < 0.01$

# Hoxb5 marks long-term haematopoietic stem cells and reveals a homogenous perivascular niche

James Y. Chen<sup>1,2\*</sup>, Masanori Miyanishi<sup>1,2\*</sup>, Sean K. Wang<sup>1,2</sup>, Satoshi Yamazaki<sup>3</sup>, Rahul Sinha<sup>1,2</sup>, Kevin S. Kao<sup>1,2</sup>, Jun Seita<sup>1</sup>, Debashis Sahoo<sup>1,2†</sup>, Hiromitsu Nakauchi<sup>1,3</sup> & Irving L. Weissman<sup>1,2</sup>

**Haematopoietic stem cells (HSCs) are arguably the most extensively characterized tissue stem cells. Since the identification of HSCs by prospective isolation<sup>1</sup>, complex multi-parameter flow cytometric isolation of phenotypic subsets has facilitated studies on many aspects of HSC biology, including self-renewal<sup>2–4</sup>, differentiation, ageing, niche<sup>5</sup>, and diversity<sup>6–8</sup>. Here we demonstrate by unbiased multi-step screening, identification of a single gene, homeobox B5 (*Hoxb5*, also known as *Hox-2.1*), with expression in the bone marrow that is limited to long-term (LT)-HSCs in mice. Using a mouse single-colour tri-mCherry reporter driven by endogenous *Hoxb5* regulation, we show that only the *Hoxb5*<sup>+</sup> HSCs exhibit long-term reconstitution capacity after transplantation in primary transplant recipients and, notably, in secondary recipients. Only 7–35% of various previously defined immunophenotypic HSCs are LT-HSCs. Finally, by *in situ* imaging of mouse bone marrow, we show that >94% of LT-HSCs (*Hoxb5*<sup>+</sup>) are directly attached to VE-cadherin<sup>+</sup> cells, implicating the perivascular space as a near-homogenous location of LT-HSCs.**

Prospective isolation of HSCs requires that the isolated cells are capable of long-term production of all blood cell types in primary irradiated hosts, as well as self-renewal, such that the cells can be transplanted to secondary hosts to give rise to long-term multilineage repopulation. From the first enrichment and isolation of candidate HSCs<sup>1,9,10</sup>, this activity has been entirely contained in cell-surface-marker-defined cell populations, and more recently in fluorescent reporters<sup>11–13</sup>. However, the precise fraction of cells in those populations that are true LT-HSCs remains unknown.

To enable further purification of LT-HSCs, we sought to identify genes expressed exclusively in HSCs within cell populations resident in mouse bone marrow, detectable by flow cytometry and *in situ* fluorescence, and thus performed the following four-step screening (Fig. 1d).

First, we compared microarray gene expression assays among 28 distinct populations of the haematopoietic system (Extended Data Fig. 1a and Supplementary Table 1). Using the Gene Expression Commons platform<sup>14</sup>, we identified 118 candidate HSC-specific genes (Fig. 1a and Supplementary Table 2). Surprisingly, this list did not include all previously reported HSC-specific markers<sup>11–13</sup> (Extended Data Fig. 1b and Supplementary Table 2). Second, to identify HSCs *in situ*, we excluded candidates that also label non-haematopoietic cells present in the bone marrow such as stromal and endothelial cells<sup>15,16</sup>. Consequently, we excluded genes expressed in eight distinct non-haematopoietic bone marrow populations, thereby narrowing the list to 45 candidate genes (Fig. 1a).

Next, to ensure that the expression of any candidates could be detected by both flow cytometry and *in situ* fluorescence, we used RNA-sequencing (RNA-seq) combined with a threshold gene standard to estimate the fragments per kilobase of transcript per million mapped reads

(FPKM) value that could serve as a detection threshold. From the bone marrow of 12-week-old mice, we sorted and RNA-sequenced immunophenotypically defined (Lin<sup>−</sup>c-Kit<sup>+</sup>Sca-1<sup>+</sup>CD150<sup>+</sup>CD34<sup>−/lo</sup>Flk2<sup>−</sup>) HSCs (hereafter referred to as pHSCs), multipotent progenitors subset A (MPPa; Lin<sup>−</sup>c-Kit<sup>+</sup>Sca-1<sup>+</sup>CD150<sup>+</sup>CD34<sup>+</sup>Flk2<sup>−</sup>), and multipotent progenitors subset B (MPPb; Lin<sup>−</sup>c-Kit<sup>+</sup>Sca-1<sup>+</sup>CD150<sup>−</sup>CD34<sup>+</sup>Flk2<sup>−</sup>) (Fig. 1b) to determine the FPKM value of candidate genes. On the basis of Bmi-1-eGFP knock-in reporter expression<sup>17</sup>, we found that a single copy of eGFP is detectable at an estimated FPKM value of ~20. However, this high threshold would have excluded all candidate genes. Therefore, we designed a targeting construct (Fig. 1e) with three copies of mCherry, bringing the theoretical detection limit to ~7 FPKM. Lastly, to minimize aberrant detection, we set threshold FPKM values for both the MPPa and MPPb fractions to 2.5. Only three genes, *Hoxb5*, *Rnf208*, and *Smtnl1*, met these criteria (Fig. 1b).

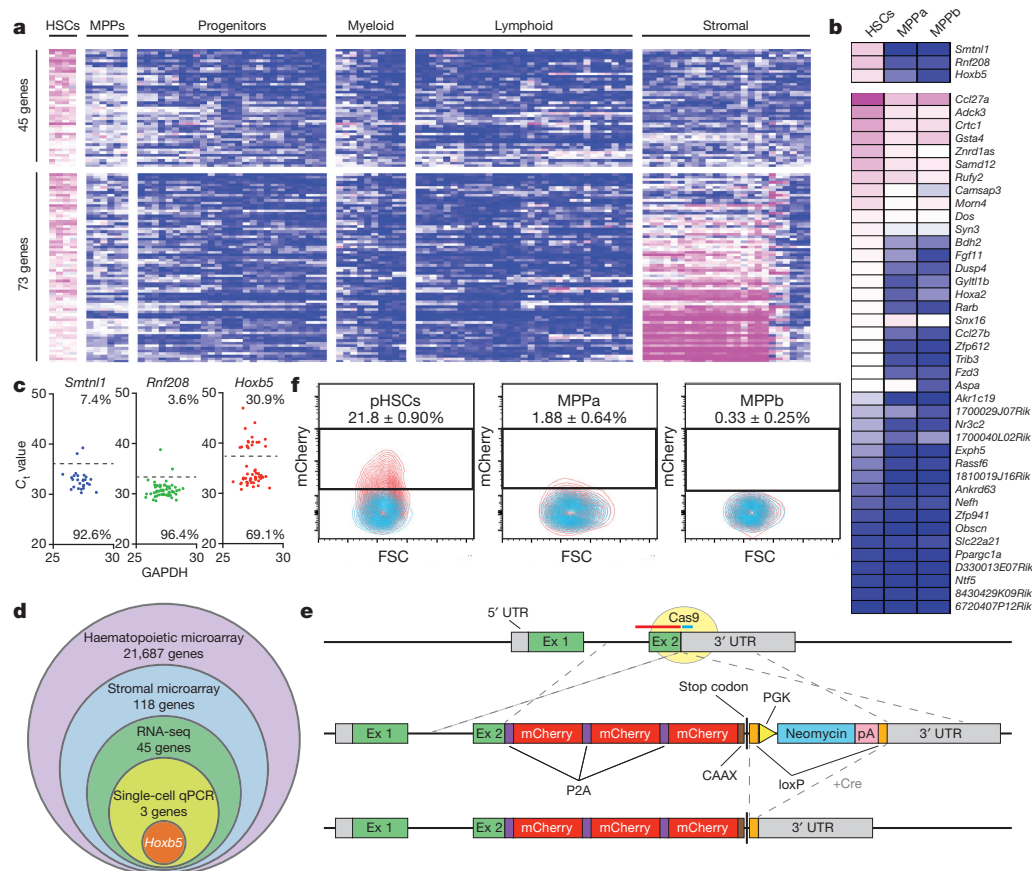
Using previous reports of heterogeneity within pHSCs<sup>7,18–20</sup>, we analysed single cells to determine whether the remaining candidate genes were heterogeneously expressed among pHSCs. We reasoned that an ideal pan-HSC candidate gene would label the majority of pHSCs, with quantitative differences potentially reflecting HSC heterogeneity/diversity. We thus performed single-cell quantitative PCR (qPCR) analysis of pHSCs, and evaluated expression of *Hoxb5*, *Rnf208*, and *Smtnl1*. Only *Hoxb5* exhibited bimodal expression, in comparison to the unimodality of *Rnf208* and *Smtnl1* (Fig. 1c). Therefore, from the entire HSC transcriptome, only *Hoxb5* satisfied the criteria of our extensive unbiased screening (Fig. 1d).

We next sought to generate a *Hoxb5* reporter with minimal disruption of endogenous *Hoxb5* function. Thus we designed our targeting construct and CRISPR guide sequences to facilitate an in-frame knock-in to the endogenous *Hoxb5* gene locus immediately 5' of the only endogenous stop codon. We used three tandem mCherry cassettes separated by porcine teschovirus-1 2A (P2A) sequences, with the terminal mCherry carrying a CAAX membrane localization sequence (*Hoxb5*-tri-mCherry) (Fig. 1e).

To evaluate the specificity of this reporter, we isolated whole bone marrow cells from 12-week-old reporter mice and measured mCherry<sup>+</sup> cells in the following immunophenotypic populations: pHSC, MPPa, MPPb, Flk2<sup>+</sup> multipotent progenitor, megakaryocyte erythrocyte progenitor, granulocyte monocyte progenitor, common myeloid progenitor, common lymphoid progenitor fractions; differentiated cell populations (B cell, T cell, natural killer (NK) cell, neutrophil, eosinophil, monocyte, macrophage, dendritic cell, red blood cell, and megakaryocyte); and in CD45<sup>−</sup> stromal fractions (Fig. 1f, Extended Data Fig. 2a, b, Extended Data Fig. 3, and unpublished data). Consistent with our initial screen (Fig. 1a–d), and using wild-type mice as a fluorescence minus one (FMO) threshold<sup>21</sup>, mCherry-labelled cells were highly enriched in the pHSC fraction (21.8% ± 0.90%), had a low

<sup>1</sup>Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, California 94305, USA. <sup>2</sup>Ludwig Center for Cancer Stem Cell Research and Medicine, Stanford University School of Medicine, Stanford, California 94305, USA. <sup>3</sup>Division of Stem Cell Therapy, Center for Stem Cell Biology and Regenerative Medicine, The Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan. †Present address: Department of Pediatrics and Department of Computer Science and Engineering, University of California San Diego, San Diego, California 92123, USA.

\*These authors contributed equally to this work.



**Figure 1 | Multi-step unbiased screening identifies *Hoxb5* as a candidate LT-HSC marker.** **a**, Microarray heat map depicting relative expression (pink, high; blue, low) of HSC-specific genes in haematopoietic and stromal populations. Each row represents a gene; each subcolumn a replicate microarray; each labelled column a category of cell populations. The 45 genes in the top panel displayed limited activity in all non-HSC populations examined. **b**, Transcriptional profiling by RNA-seq of the 45 genes from **a**. Three genes (top panel) exceeded the estimated threshold for detection (FPKM > 7.0) in HSCs while showing minimal expression

frequency in the MPPa fraction ( $1.88\% \pm 0.64\%$ ), and background frequencies in the remaining fractions (Fig. 1f, Extended Data Figs 2b and 3 and unpublished data). Interestingly, as only a minority of pHSCs were mCherry<sup>+</sup>, this suggested that either our reporter labelled only a subfraction of HSCs or only a subfraction of pHSCs were indeed HSCs.

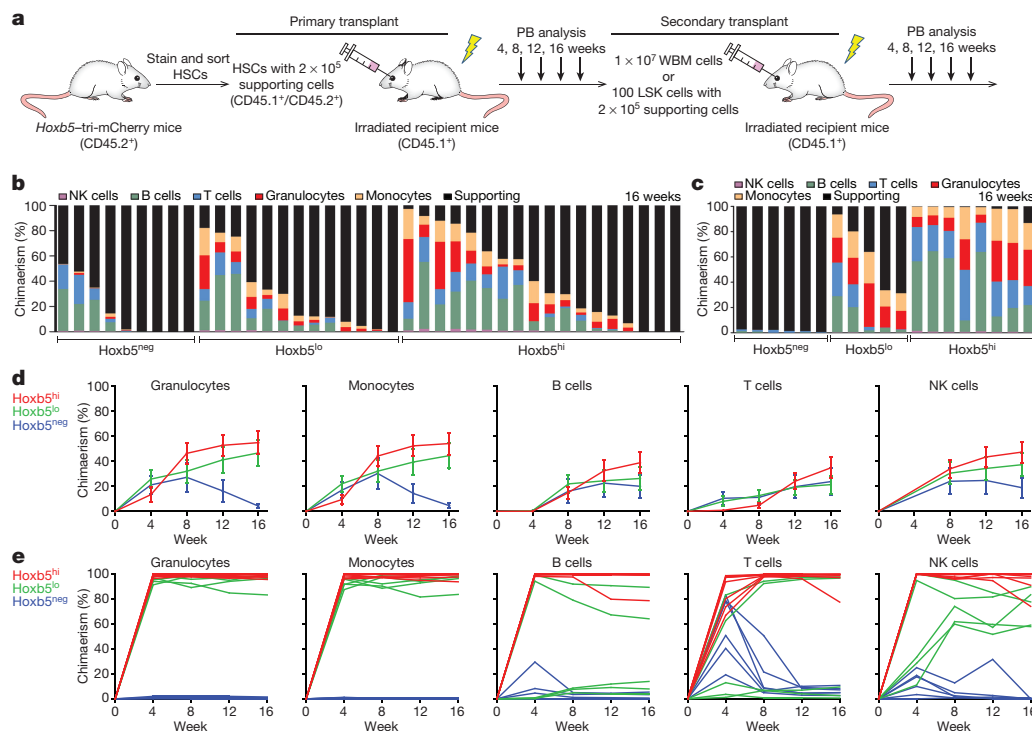
To distinguish between these two possibilities and to determine whether *Hoxb5* is a reporter of LT-HSCs, we characterized *Hoxb5*-expressing cells by transplantation. In order to be inclusive of all events in the pHSC gate, we used wild-type FMO to define *Hoxb5* negativity (*Hoxb5*<sup>neg</sup> or sometimes referred to as *Hox5*<sup>-</sup>) and divided the positive fraction (sometimes referred to as *Hox5*<sup>+</sup>) into *Hoxb5*<sup>hi</sup> (top 5<sup>th</sup> percentile) and *Hoxb5*<sup>lo</sup> ( $14.1 \pm 7.7\%$ ) (Extended Data Fig. 4). Ten-cell and three-cell grafts of *Hoxb5*<sup>hi</sup>, *Hoxb5*<sup>lo</sup>, or *Hoxb5*<sup>neg</sup> pHSCs were transplanted with supporting bone marrow cells into irradiated mouse recipients. We used CD45.2 expression to assess donor HSC contribution to haematopoietic lineages at 4-week intervals (Fig. 2a). Analysis of peripheral blood 16 weeks after transplantation of ten cells demonstrated that multilineage reconstitution was present in 78% of *Hoxb5*<sup>hi</sup>, 70% of *Hoxb5*<sup>lo</sup>, and 44% *Hoxb5*<sup>neg</sup> pHSC recipients (Fig. 2b and Extended Data Fig. 5a). Three-cell transplants exhibited similar kinetics (Extended Data Fig. 5c). Notably, the chimaerism of the *Hoxb5*<sup>neg</sup> pHSCs decreased over time, in particular between 4 to 8 weeks and shifted towards a predominantly lymphoid chimaerism (Fig. 2b, d and Extended Data Fig. 5a), suggesting that either the

(FPKM < 2.5) in MPPa and MPPb populations. **c**, Heterogeneity of expression for the three remaining candidate genes in HSCs as assessed by single-cell qPCR. **d**, Venn diagram reflecting the four-step screen that identified *Hoxb5* as an ideal candidate in the HSC transcriptome. **e**, Targeting strategy to generate a triple-mCherry *Hoxb5* knock-in mouse reporter line (*Hoxb5*-tri-mCherry). UTR, untranslated region; PGK, phosphoglycerate kinase I. **f**, *Hoxb5* reporter expression (red) in pHSCs and MPPs compared to wild-type controls (blue). Values indicate the percentage of mCherry<sup>+</sup> cells  $\pm$  s.d. in each fraction for  $n = 3$  mice.

*Hoxb5*<sup>neg</sup> fraction comprised of lymphoid-biased HSCs or, more likely, transiently self-renewing short-term (ST)-HSCs and/or MPPs that had given rise to long-lived lymphocytes.

To evaluate these two possibilities, we carried out a secondary transplantation of whole bone marrow from primary *Hoxb5*<sup>hi</sup>, *Hoxb5*<sup>lo</sup>, or *Hoxb5*<sup>neg</sup> pHSC recipients into lethally irradiated secondary recipients (Fig. 2a). Sixteen weeks after secondary transplantation, peripheral blood analyses revealed robust multilineage chimaerism from all *Hoxb5*<sup>hi</sup> and *Hoxb5*<sup>lo</sup> transplant recipients, with minimal chimaerism from the *Hoxb5*<sup>neg</sup> fraction (Fig. 2c, e and Extended Data Fig. 5b). Furthermore, bone marrow analysis of primary recipients for donor pHSCs revealed that the *Hoxb5*<sup>hi</sup> pHSC recipients contained *Hoxb5*<sup>hi</sup>, *Hoxb5*<sup>lo</sup>, and *Hoxb5*<sup>neg</sup> cells (100%,  $n = 10$  mice), whereas the recipients of *Hoxb5*<sup>neg</sup> cells were only *Hoxb5*<sup>neg</sup> (36%,  $n = 4$  mice) or devoid of donor cells (64%,  $n = 7$  mice). These results further suggested that the *Hoxb5*<sup>neg</sup> pHSCs were in fact transiently self-renewing ST-HSCs/MPPs. To determine if *Hoxb5* expression could also distinguish LT-HSCs in a primary transplant, we normalized the number of donor cells used for secondary transplant by sorting 100 Lin<sup>-</sup>Kit<sup>+</sup>Sca-1<sup>+</sup> (LSK) donor cells from the primary *Hoxb5*<sup>hi</sup> recipients and transplanted them into irradiated secondary recipients ( $n = 24$  mice) (Fig. 2a). As with the primary bone marrow transplantations, the chimaerism was minimal in *Hoxb5*<sup>-</sup> compared to *Hoxb5*<sup>+</sup> recipients (Extended Data Fig. 5d). Limiting dilution analysis revealed that the frequency of long-term plus short-term HSCs in primary





**Figure 2 | *Hoxb5* distinguishes between LT-HSC and non-LT-HSC.** **a**, Experimental schematic for long-term haematopoietic reconstitution assays. CD45.1<sup>+</sup> recipient mice were lethally irradiated and competitively transplanted with ten or three *Hoxb5*<sup>tri-mCherry</sup> HSCs and  $2 \times 10^5$  CD45.1<sup>+</sup>/CD45.2<sup>+</sup> supporting cells. For secondary transplants,  $1 \times 10^7$  whole bone marrow (WBM) cells or 100 sorted LSK cells were transferred from primary recipient mice. PB, peripheral blood. **b**, Percentage chimaerism at 16 weeks in primary recipients receiving ten *Hoxb5*<sup>neg</sup> ( $n = 9$  mice), *Hoxb5*<sup>lo</sup> ( $n = 13$  mice), or *Hoxb5*<sup>hi</sup> ( $n = 18$  mice) pHSCs. Each column represents an individual mouse. **c**, Percentage chimaerism at 16 weeks following whole bone marrow secondary transplant. **d**, Average donor lineage contribution in ten-cell primary transplants. **e**, Individual donor chimaerism by lineage in whole bone marrow secondary recipients. Each line represents an individual mouse ( $n = 6$  mice for *Hoxb5*<sup>neg</sup>;  $n = 5$  mice for *Hoxb5*<sup>lo</sup>; and  $n = 8$  mice for *Hoxb5*<sup>hi</sup>).

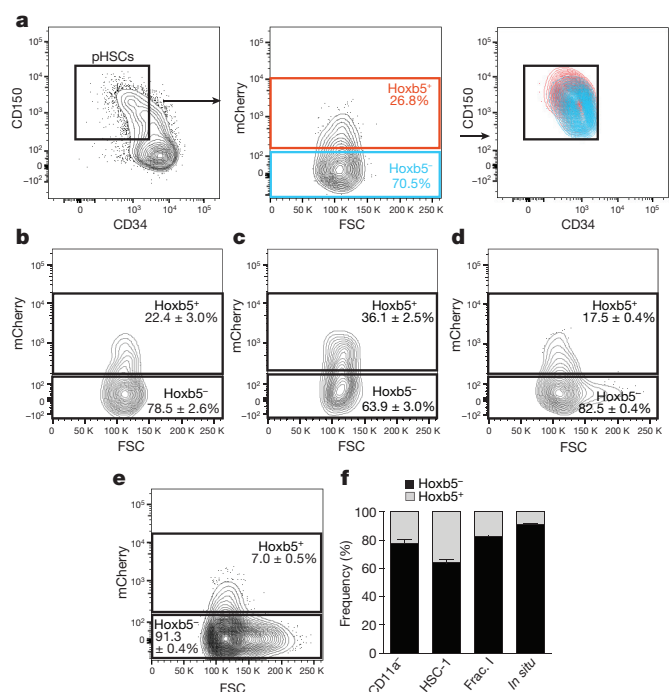
hosts at 16 weeks was 1 in 2.1 for *Hoxb5*<sup>hi</sup>, 1 in 2.4 for *Hoxb5*<sup>lo</sup>, and 1 in 16.1 for *Hoxb5*<sup>neg</sup> recipients (Extended Data Fig. 6). Taken together, these results demonstrate that *Hoxb5* labels functional LT-HSCs.

Given that *Hoxb5*<sup>+</sup> cells are non-LT-HSCs, we re-examined the specificity of past definitions (Fig. 3a, f), including previously reported refinements to the LT-HSC immunophenotype<sup>18–20</sup> and the most widely used *in situ* definition over the past decade<sup>22</sup>. We found that  $78.5 \pm 2.6\%$  of CD11a<sup>+</sup> HSCs<sup>19</sup>,  $63.9 \pm 3\%$  of the HSC-1 (Lin<sup>−</sup>c-Kit<sup>+</sup>Sca-1<sup>+</sup>CD48<sup>−</sup>CD150<sup>+</sup>CD229<sup>−</sup>CD244<sup>−</sup>)<sup>20</sup>, and  $82.5 \pm 0.4\%$  of fraction I HSCs (Lin<sup>−</sup>c-Kit<sup>+</sup>Sca-1<sup>+</sup>CD34<sup>−/lo</sup>CD150<sup>+</sup>CD41<sup>−</sup>)<sup>18</sup> were *Hoxb5*<sup>+</sup> (Fig. 3b, c, and d, respectively). Surprisingly,  $91.3 \pm 0.4\%$  of Lin<sup>−</sup>CD48<sup>−</sup>CD41<sup>−</sup>CD150<sup>+</sup> (ref. 22) cells were *Hoxb5*<sup>+</sup> (Fig. 3e). As this subset was initially used for localization of HSCs *in situ*<sup>22</sup>, we re-examined the *in situ* location of HSCs using *Hoxb5* expression.

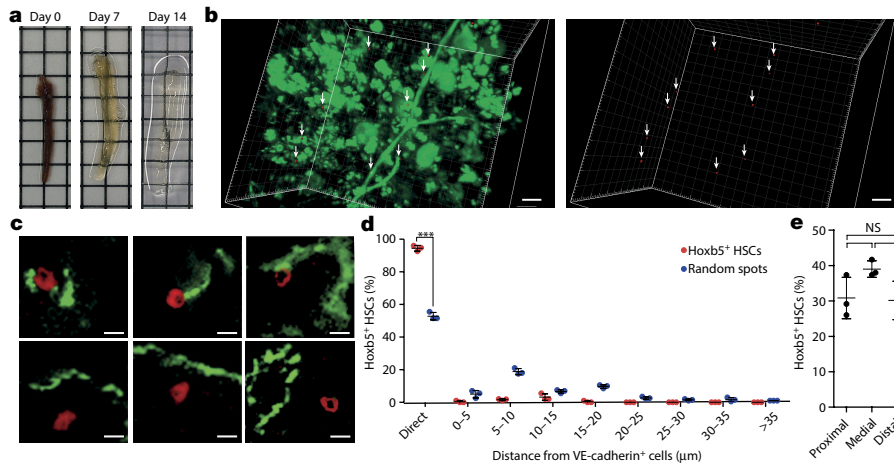
Visualizing LT-HSCs in bone marrow and identifying the cellular constituents and structures of the HSC niche remains challenging. Despite this, multiple constituent cell types have been proposed, including mesenchymal stromata, endosteal osteoblasts, glia, endothelia, and pericytes<sup>5</sup>. *In situ* studies are made difficult by several technical limitations, including the number of fluorescent colours, difficulty in identifying HSCs surrounded by non-HSCs, and difficulty in translating the same fluorescently defined positive and negative thresholds from flow cytometry to tissue sections.

To address these issues, we used the *Hoxb5*<sup>tri-mCherry</sup> reporter. Using flow cytometry, we determined the utility of *Hoxb5* alone in identifying LT-HSCs. After logical exclusion of autofluorescence by comparison to wild-type control mice, we found that all *Hoxb5*<sup>+</sup> events are within the c-Kit<sup>+</sup> fraction and  $62.0 \pm 12.8\%$  of all *Hoxb5*<sup>+</sup> events are located in the pHSC gate (Extended Data Fig. 8), representing an eight- to nine-fold enrichment compared to previous *in situ* labelling of HSCs (Lin<sup>−</sup>CD48<sup>−</sup>CD41<sup>−</sup>CD150<sup>+</sup>)<sup>22</sup> (Fig. 3e).

To reveal the three-dimensional (3D) HSC niche architecture, we applied the CUBIC technique (clear, unobstructed brain imaging cocktails and computational analysis<sup>23</sup>) to tibial bone marrow, facilitating depletion of autofluorescent cellular components (Fig. 4a). Given



**Figure 3 | Previously defined HSC immunophenotypes contain *Hoxb5*<sup>+</sup> cells.** Flow cytometry plots of bone marrow from 12-week-old mice depicting *Hoxb5*<sup>tri-mCherry</sup> reporter activity in previously reported HSC immunophenotypes. **a**, pHSC (LSK CD150<sup>+</sup>CD34<sup>−/lo</sup>Flk2<sup>−</sup>) *Hoxb5*<sup>+</sup> (red) and *Hoxb5*<sup>−</sup> (blue). **b**, CD11a<sup>+</sup> (LSK CD150<sup>+</sup>CD34<sup>−/lo</sup>CD11a<sup>+</sup>)<sup>19</sup>. **c**, HSC-1 (LSK CD150<sup>+</sup>CD48<sup>−/lo</sup>CD229<sup>−/lo</sup>CD244<sup>−</sup>)<sup>20</sup>. **d**, Fraction I (LSK CD150<sup>+</sup>CD34<sup>−/lo</sup>CD41<sup>−</sup>)<sup>18</sup>. **e**, CD150<sup>+</sup>CD48<sup>−</sup>CD41<sup>−</sup> cells, currently used to identify HSCs *in situ*<sup>22</sup>. Wild-type FMO used to define *Hoxb5* negativity for each panel. **f**, Summary percentage of *Hoxb5*<sup>+</sup> and *Hoxb5*<sup>−</sup> negative cells in characterized HSC subfractions. Error bars denote s.d.;  $n = 5$  mice (see Extended Data Fig. 7a–d for gating scheme).



**Figure 4 | LT-HSCs exhibit near-homogenous attachment to VE-cadherin<sup>+</sup> cells.** **a**, Tissue preparation and representative images of tibial bone marrow plug after paraformaldehyde fixation (day 0) and treatment with reagent-1 (see ref. 23) (day 7, day 14). **b**, Localization of Hoxb5<sup>+</sup> cells (red and arrows) and VE-cadherin<sup>+</sup> cells (green) in 3D-reconstructed images. Scale bar, 30 μm. **c**, Representative 2D images of direct (top panel) and non-direct (bottom panel) association of

Hoxb5<sup>+</sup> cells (red) with VE-cadherin<sup>+</sup> cells (green). Scale bar, 10 μm. **d**, Frequency of Hoxb5<sup>+</sup> cells ( $n = 287$  cells, from  $n = 3$  mice) and random spots ( $n = 600$  spots, from  $n = 3$  mice) plotted against proximity to VE-cadherin<sup>+</sup> cells. \*\*\* $P < 0.0001$ . **e**, Average number of Hoxb5<sup>+</sup> cells in proximal, medial, and distal regions of tibia ( $n = 3$  mice). NS, not significant. Unpaired Student's *t*-test (**d**, **e**).

that fluorescence intensity is preserved by CUBIC<sup>23</sup>, the specificity of LT-HSCs *in situ* correlates with Hoxb5 intensity determined by flow cytometry. Therefore, on the basis of the number of Hoxb5<sup>+</sup> LT-HSCs detected in one tibia (Extended Data Fig. 9a, b), we objectively gated the same number of Hoxb5<sup>+</sup> cells by intensity as *in situ* Hoxb5<sup>+</sup> cells. We observed a uniform distribution of *in situ* Hoxb5<sup>+</sup> cells along the longitudinal axis of the tibia (Fig. 4e and Extended Data Fig. 10a–c). The average percentage of *in situ* Hoxb5<sup>+</sup> cells was  $30.8 \pm 5.9$  in the proximal epiphysis,  $39.0 \pm 2.4$  in the diaphysis, and  $30.1 \pm 5.4$  in the distal epiphysis per field ( $1.25 \times 1.25 \times 0.4 \text{ mm}^3$ ) (Fig. 4e).

To investigate the association of LT-HSCs with vasculature, we injected anti-mouse VE-cadherin antibody into reporter mice and analysed the bone marrow with CUBIC<sup>23</sup>. We found that  $94.1 \pm 1.9\%$  of *in situ* Hoxb5<sup>+</sup> cells were directly attached to the abluminal surface of VE-cadherin<sup>+</sup> endothelial cells. In contrast, only  $52.8 \pm 2.3\%$  of the random (Hoxb5<sup>−</sup>) spots were directly associated with VE-cadherin<sup>+</sup> cells. This implied a near-homogenous perivascular location for the LT-HSC niche (Fig. 4b–d).

Following the first successful enrichment of HSCs in 1988 (ref. 1), many groups have attempted to identify surface markers to isolate LT-HSCs<sup>18–20,22,24,25</sup>. Identification of CD150, CD34, and CD48 enabled isolation of LT/ST-HSCs from MPPs. However, complete separation of LT-HSCs from ST-HSCs has never been fully accomplished. Our study demonstrates that *Hoxb5* expression is specifically limited to LT-HSCs in adult mouse bone marrow. Limiting dilution assay shows that at least 1 in 2.1 of Hoxb5<sup>hi</sup> cells are LT-HSCs. However, this assay underestimates the functional potential of candidate HSCs. Transplantation, although the gold standard for prospective isolation of HSCs, may not take into account differences between sessile and mobile HSCs<sup>26,27</sup>, cell cycle status<sup>28,29</sup>, expression of CD47 (an anti-phagocytic molecule expressed highly on mobilized HSCs but at low levels on sessile bone marrow HSCs<sup>30</sup>), irradiation, and their influence on engraftment efficiency.

Our results demonstrate that the LT-HSCs compartment is near-homogeneously perivascular. Although other compartments have been implicated<sup>5</sup>, including less homogenous association of candidate HSCs with the vasculature<sup>11</sup>, it is likely that our results differ owing to the fraction of assayed populations that are LT-HSCs (Extended Data Fig. 1b and Supplementary Table 3). These results suggest that the model of *Hoxb5*-tri-mCherry is the most specific available, to our knowledge, for identification of LT-HSCs, and could facilitate *in situ*

lineage tracing to define their role in haematopoiesis in the absence of transplantation into irradiated hosts.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 16 June; accepted 21 December 2015.**

- Spangrude, G. J., Heimfeld, S. & Weissman, I. L. Purification and characterization of mouse hematopoietic stem cells. *Science* **241**, 58–62 (1988).
- Guo, W. *et al.* Multi-genetic events collaboratively contribute to *Pten*-null leukaemia stem-cell formation. *Nature* **453**, 529–533 (2008).
- Yilmaz, O. H. *et al.* *Pten* dependence distinguishes haematopoietic stem cells from leukaemia-initiating cells. *Nature* **441**, 475–482 (2006).
- Ito, K. *et al.* Regulation of oxidative stress by ATM is required for self-renewal of haematopoietic stem cells. *Nature* **431**, 997–1002 (2004).
- Morrison, S. J. & Scadden, D. T. The bone marrow niche for haematopoietic stem cells. *Nature* **505**, 327–334 (2014).
- Dykstra, B. *et al.* Long-term propagation of distinct hematopoietic differentiation programs *in vivo*. *Cell Stem Cell* **1**, 218–229 (2007).
- Beeraman, I. *et al.* Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion. *Proc. Natl Acad. Sci. USA* **107**, 5465–5470 (2010).
- Lu, R., Neff, N. F., Quake, S. R. & Weissman, I. L. Tracking single hematopoietic stem cells *in vivo* using high-throughput sequencing in conjunction with viral genetic barcoding. *Nature Biotechnol.* **29**, 928–933 (2011).
- Uchida, N. & Weissman, I. L. Searching for hematopoietic stem cells: evidence that Thy-1.1<sup>lo</sup> Lin<sup>−</sup> Sca-1<sup>+</sup> cells are the only stem cells in C57BL/Ka-Thy-1.1 bone marrow. *J. Exp. Med.* **175**, 175–184 (1992).
- Morrison, S. J. & Weissman, I. L. The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity* **1**, 661–673 (1994).
- Acar, M. *et al.* Deep imaging of bone marrow shows non-dividing stem cells are mainly perisinusoidal. *Nature* **526**, 126–130 (2015).
- Gazit, R. *et al.* *Fgd5* identifies hematopoietic stem cells in the murine bone marrow. *J. Exp. Med.* **211**, 1315–1331 (2014).
- Hills, D. *et al.* Hoxb4-YFP reporter mouse model: a novel tool for tracking HSC development and studying the role of Hoxb4 in hematopoiesis. *Blood* **117**, 3521–3528 (2011).
- Seita, J. *et al.* Gene expression commons: an open platform for absolute gene expression profiling. *PLoS ONE* **7**, e40321 (2012).
- Chan, C. K. F. *et al.* Identification and specification of the mouse skeletal stem cell. *Cell* **160**, 285–298 (2015).
- Chan, C. K. F. *et al.* Clonal precursor of bone, cartilage, and hematopoietic niche stromal cells. *Proc. Natl Acad. Sci. USA* **110**, 12643–12648 (2013).
- Hosen, N. *et al.* Bmi-1-green fluorescent protein-knock-in mice reveal the dynamic regulation of bmi-1 expression in normal and leukemic hematopoietic cells. *Stem Cells* **25**, 1635–1644 (2007).

18. Yamamoto, R. *et al.* Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. *Cell* **154**, 1112–1126 (2013).
19. Fathman, J. W. *et al.* Upregulation of CD11A on hematopoietic stem cells denotes the loss of long-term reconstitution potential. *Stem Cell Reports* **3**, 707–715 (2014).
20. Oguro, H., Ding, L. & Morrison, S. J. SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. *Cell Stem Cell* **13**, 102–116 (2013).
21. Herzenberg, L. A., Tung, J., Moore, W. A., Herzenberg, L. A. & Parks, D. R. Interpreting flow cytometry data: a guide for the perplexed. *Nature Immunol.* **7**, 681–685 (2006).
22. Kiel, M. J. *et al.* SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**, 1109–1121 (2005).
23. Susaki, E. A. *et al.* Whole-brain imaging with single-cell resolution using chemical cocktails and computational analysis. *Cell* **157**, 726–739 (2014).
24. Christensen, J. L. & Weissman, I. L. Flk-2 is a marker in hematopoietic stem cell differentiation: a simple method to isolate long-term stem cells. *Proc. Natl Acad. Sci. USA* **98**, 14541–14546 (2001).
25. Osawa, M., Hanada, K., Hamada, H. & Nakauchi, H. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science* **273**, 242–245 (1996).
26. Wright, D. E., Wagers, A. J., Gulati, A. P., Johnson, F. L. & Weissman, I. L. Physiological migration of hematopoietic stem and progenitor cells. *Science* **294**, 1933–1936 (2001).
27. Forsberg, E. C. *et al.* Molecular signatures of quiescent, mobilized and leukemia-initiating hematopoietic stem cells. *PLoS ONE* **5**, e8785 (2010).
28. Fleming, W. H. *et al.* Functional heterogeneity is associated with the cell cycle status of murine hematopoietic stem cells. *J. Cell Biol.* **122**, 897–902 (1993).
29. Passegué, E., Wagers, A. J., Giuriato, S., Anderson, W. C. & Weissman, I. L. Global analysis of proliferation and cell cycle gene expression in the regulation of hematopoietic stem and progenitor cell fates. *J. Exp. Med.* **202**, 1599–1611 (2005).
30. Jaiswal, S. *et al.* CD47 is upregulated on circulating hematopoietic stem cells and leukemia cells to avoid phagocytosis. *Cell* **138**, 271–285 (2009).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We would like to acknowledge N. Neff and G. Mantalas for advice regarding RNA sequencing; B. Yu and A. Beel for providing critical input on imaging data; H. Nishikii for advice regarding imaging data; S. Karten for help in editing the manuscript; L. Jerabek and T. Storm for laboratory management; A. McCarty and C. Wang for animal care; P. Lovelace and J. Ho for FACS facility management; H. Zeng, Y. Li, and C. Wang for collaboration in generating the mouse model; and Y. Sato for technical assistance in Imaris software analysis. The authors would like to acknowledge ongoing support for this work: NCI and NHLBI of the NIH under award numbers R01 CA086065, U01 HL099999, and R01 HL058770, and by the Virginia and D. K. Ludwig Fund for Cancer Research (I.L.W.); Stanford University Medical Scientist Training Program (T32 GM007365) and NHLBI Ruth L. Kirschstein National Research Service Award (F30-HL122096) (J.Y.C.); and Human Frontier Science Program Long-Term Fellowships, the Uehara Memorial Foundation Research Fellowship, Toyobo Biotechnology Foundation Research Fellowship, and Kanzawa Medical Research Foundation Overseas study grants (M.M.). The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH.

**Author Contributions** J.Y.C. and M.M. contributed equally to this work, and either has the right to list himself first in bibliographic documents. M.M. and J.Y.C. conceived, performed, analysed, and oversaw the experiments, with suggestions from I.L.W. M.M. and J.Y.C. identified *Hoxb5* as a LT-HSC marker, and made and characterized the *Hoxb5*-tri-mCherry mouse. S.K.W. and K.S.K. performed experiments and prepared figures under the supervision of M.M. and J.Y.C. S.Y. generated CUBIC data and evaluated the association with VE-cadherin vasculature. R.S. designed and performed RNA-seq and associated data analysis. J.S. and D.S. designed the gene expression commons for microarray analysis. D.S. provided critical advice regarding combined analysis of microarray and RNA-seq data. M.M., J.Y.C., S.K.W., K.S.K., and I.L.W. wrote the manuscript. H.N. and R.S. provided comments on the manuscript.

**Author Information** Microarray data was deposited at GEO under accession number GSE77078. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.M. ([supamasa@stanford.edu](mailto:supamasa@stanford.edu)) or I.L.W. ([irv@stanford.edu](mailto:irv@stanford.edu)).



## METHODS

**Data Reporting.** No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to outcome assessment.

**Mice.** Eight-to-twelve-week-old C57BL/6J male mice (Jackson Laboratory) were used as wild-type controls. Eight-to-twelve-week-old male B6.SJL-*Ptprca*<sup>a</sup>*Pepec*<sup>b</sup>/BoyJ mice (Jackson Laboratory) were used as recipients for transplantation assay. Supporting cells for competitive reconstitution assays were collected from B6.SJL-*Ptprca*<sup>a</sup>*Pepec*<sup>b</sup>/BoyJ × C57BL/6J (F<sub>1</sub> mice CD45.1<sup>+</sup>/CD45.2<sup>+</sup>). Mice were bred at our animal facility according to NIH guidelines. All animal protocols were approved by the Stanford University Administrative Panel on Laboratory Animal Care. *Hoxb5*-tri-mCherry (C57BL/6J background) mice were used as donor cells for transplantation as well as for analysis. Please see Gene targeting for mouse derivation.

**Microarray data.** All microarray data employed in this study are available at Gene Expression Commons (<http://gexc.stanford.edu>) and GEO GSE77078. Two to four microarray replicates were assessed for each distinct cell population<sup>14–16</sup>. The immunophenotype definition of each fraction is included in Supplementary Table 1.

**RNA sequencing.** RNA sequencing was performed as previously described<sup>31</sup>. In brief, total RNA was isolated with trizol, treated with RQ1 RNase free DNase (Promega) to remove minute quantities of genomic DNA if present, and cleaned up using RNeasy minelute columns (QIAGEN). cDNA libraries were prepared for pHSC, MPPa, and MPPb populations using Ovation RNA-Seq System V2 (NuGen) and sequenced separately using HiSeq 2500 (Illumina) to obtain 2 × 150 base pair (bp) paired-end reads. Raw transcriptome sequence data were mapped to *Mus musculus* reference mRNAs using Olego<sup>32</sup> to produce a reference-guided transcript assembly. Four replicates were sequenced for each population. Data are accessible at NCBI SRP068593.

**Single-cell qPCR.** Single cells for qPCR were processed using the Single Cell-to-CT qRT-PCR Kit (Life Technologies) as per the manufacturer's instructions. Cells were sorted directly into lysis solution in a 96-well plate, subjected to a reverse transcription reaction for cDNA synthesis, amplified for 14 cycles with pooled TaqMan Gene Expression Assays, and diluted with 1 × TE buffer (pH 8.0). For real-time PCR, samples were amplified for 50 cycles of 3 s at 95 °C followed by 30 s at 60 °C using the following TaqMan probes: Mm00657672\_m1 for *Hoxb5*; Mm03039759\_s1 for *Rnf208*; Mm00470338\_m1 for *Smtnl1*; and Mm99999915\_g1 for *Gapdh*. All thermocycling was performed on a 7900HT Fast Real-Time PCR System (Applied Biosystems). Presence of a single cell was validated by a C<sub>t</sub> value of 30 or less for *Gapdh*.

**Gene targeting.** The targeting construct was cloned into pJYC (derived from pUC19, golden gate insert of TALE backbone vector<sup>33</sup>, and removal of LacZ sequence). Right and left homology arms of ~700 bp were cloned by PCR from C57BL/6J genomic DNA, and the entire construct was sequence validated. BRUCE-4 embryonic stem (ES) cells (Millipore) derived from C57BL/6J mice were transfected with the targeting construct as well as bicistronic CRISPR vector PX330 expressing Cas9 and a *Hoxb5*-specific single guide RNA (5'-GGCUCUCCGGAUGGGCUCA-3')<sup>34</sup>. Following homologous recombination, recombinant clones were positively selected for one week with neomycin (100 µg ml<sup>-1</sup>), transfected with an EF1α-Cre-puromycin vector, and selected for a second week with puromycin (1 µg ml<sup>-1</sup>). Serially diluted ES colonies were then individually picked, expanded, and screened by PCR and qPCR for site-specific integration, exclusion of off-target effects, and correct copy number. After sequence validation of the targeted site, successfully targeted ES clones were used to establish chimaeras. Chimaeric mice were crossed with C57BL/6 - *Tyr<sup>cr-2j</sup>/J* female mice to establish germline transmission.

**Flow cytometry and cell sorting.** Flow cytometry and cell sorting were performed on a FACS Aria II cell sorter (BD Biosciences) and analysed using FlowJo software (Tree Star). Bone marrow cells were collected from bilateral tibias, femurs, humeri, and pelvises by crushing (unless otherwise specified) using a mortar and pestle in Ca<sup>2+</sup>- and Mg<sup>2+</sup>-free PBS supplemented with 2% heat-inactivated bovine serum (Gibco) and 2 mM EDTA. Cells were passed through 100 µm, 70 µm, and 40 µm strainers before analysis and sorting. To enrich HSCs and progenitor populations, cells were stained with APC-conjugated anti-c-Kit (2B8) and fractionated using anti-APC magnetic beads and LS columns (both Miltenyi Biotec). c-Kit<sup>+</sup> cells were then stained with combinations of antibodies against the following surface markers: Sca-1, Flk2, CD150, CD34, IL-7R, CD16/32, and the lineage markers Ter-119, B220, CD2, CD3, CD4, CD5, CD8a, Gr-1, CD11a, CD11b, CD41, CD48, CD229, and CD244. For lymphoid populations, bone marrow cells were stained with antibodies against CD3, CD4, CD8a, CD11b, CD11c, Gr-1, NK-1.1, Ter-119, and F4/80. Antibody staining was performed at 4 °C and cells were incubated for 30 min. Cells stained with CD34 were incubated for 90 min. Before sorting or analysis, cells were stained with SYTOX Red Dead Cell Stain (Life Technologies) to assess viability as per the manufacturer's recommendations. Transplanted cells were double-sorted for purity. Further details regarding flow cytometry reagents are provided in Supplementary Table 4.

**Transplantations and peripheral blood analyses.** B6.SJL-*Ptprca*<sup>a</sup>*Pepec*<sup>b</sup>/BoyJ (Jackson Laboratory) recipient mice were lethally irradiated at a single dose of 9.1 Gy. For reconstitution assays, donor cells were first combined with 2 × 10<sup>5</sup> whole bone marrow supporting cells (B6.SJL-*Ptprca*<sup>a</sup>*Pepec*<sup>b</sup>/BoyJ × C57BL/6J F<sub>1</sub> mice CD45.1<sup>+</sup>/CD45.2<sup>+</sup>) in 200 µl of PBS with 2% FBS, then injected into the retro-orbital venous plexus. Peripheral blood analyses were performed at 4, 8, 12, and 16 weeks after primary and secondary transplantations. At each time point, 50 µl of blood was collected from the tail vein and added to 100 µl of PBS with 2 mM EDTA. Red blood cells were subsequently lysed using BD Pharm Lyse Buffer (BD Pharmingen), as per the manufacturer's protocol, for 3 min on ice, followed by blocking with 5 µg ml<sup>-1</sup> rat IgG. Leukocytes were stained with antibodies against (refer to Supplementary Table 4 for specific clones and colours): CD45.1 (FITC), CD45.2 (PE), CD11b (BUV395), Gr-1 (Alexa-Fluor700), B220 (BV786), CD3 (BV421), TCRβ (BV421), and NK-1.1 (PerCP-cy5.5). For each mouse, the percentage of donor chimaerism in the peripheral blood was defined as the percentage of CD45.1<sup>+</sup>CD45.2<sup>+</sup> cells among total CD45.1<sup>+</sup>CD45.2<sup>+</sup> and CD45.1<sup>+</sup>CD45.2<sup>+</sup> cells. To control for variability in host response to lethal irradiation, mice with host chimaerism of 50% or higher 16 weeks after transplantation were excluded from our analyses. The frequency of chimaerism in peripheral blood was analysed as follows. For evaluation of donor (CD45.1<sup>+</sup>CD45.2<sup>+</sup>) chimaerism kinetics, after exclusion of recipient (CD45.1<sup>+</sup>CD45.2<sup>-</sup>) fraction, the frequency of the donor fraction was calculated. Within the whole donor fraction, the frequency of each lineage (NK cell, B cell, T cell, granulocyte, and monocyte) was determined. For evaluation of lineage contribution kinetics, after gating each lineage, the frequency of the donor fraction (CD45.1<sup>+</sup>CD45.2<sup>+</sup>) was calculated. Any recipients that exhibited lower than 1% of chimaerism were treated as negative to exclude ambiguous cases.

**Limiting dilution analysis.** The frequency of long-term and short-term HSCs was calculated using the transplantation data of ten- and three-cell *Hoxb5*<sup>hi</sup>, *Hoxb5*<sup>lo</sup>, or *Hoxb5*<sup>neg</sup> pHSC transplants. Any mice showing long-term (>16 week) multi-lineage reconstitution (>1% in each lineage) were counted as positive recipients. A nonlinear regression semi-log best fit line was used to calculate the frequency of LT/ST-HSCs at F<sub>0</sub> = 0.368 (GraphPad Prism 6).

**CUBIC bone marrow imaging.** Bone clearing protocol was modified from the original CUBIC protocol<sup>23</sup>. Specifically, tibias were collected and fixed in 4% PFA solution for two days, after which bone marrow plugs were extracted from the distal end by flushing method with a 25-gauge syringe. For nuclear staining, bone marrow plugs were immersed in DAPI/PBS solution at 37 °C for three days with gentle shaking. For clearing, marrow plugs were immersed in ScaleCUBIC-1 (Reagent-1) at 37 °C for two weeks with gentle shaking. The solution was changed every 48 h. To visualize vasculature, 20 µg of Alexa488-conjugated anti-mouse VE-cadherin antibody (BV13) was administered intravenously (retro-orbital) with tibias collected 30 min later. Processed plugs were embedded in 4-mm-diameter glass capillaries with 2% agarose for imaging. Images were acquired using a Zeiss Z1 Lightsheet microscope (Zeiss) and reconstituted into 3D images using Zen software (Zeiss). Acquired 3D images were analysed with Imaris software (Bitplane). After exclusion of outliers including events of extraordinary size (>30 µm) or intensity, all other mCherry<sup>+</sup> (*Hoxb5*<sup>+</sup>) cells were analysed in the tibia (*n* = 287 cells in total from *n* = 3 mice). The mCherry-negative threshold was determined on the basis of the intensity level of the wild-type control tibial plugs. These mCherry<sup>-</sup> events with intensities ranging from 0.002–75.998 (75.998 representing the upper bound of mCherry intensity) were transformed into an integer list of 75,000 values (75,000 representing the length of the list), and 600 spots were then randomly selected from this integer list using the 'randbetween' (1,75,000) function in Excel 2015 (Microsoft). Using the list of random spots identified by intensity, the location and distance to VE-cadherin<sup>+</sup> cells was measured using the Imaris software. All CUBIC imaging experiments were performed in biological triplicates from three mice.

**Genotyping.** Genomic DNA from *Hoxb5*-tri-mCherry mice was isolated from tail biopsies using QuickExtract solution (EpiCentre). PCR amplification was performed using the same forward primer (5'-GACGTATCGAGATCGC CCAC-3') with two reverse primers to distinguish between the *Hoxb5*-tri-mCherry (5'-CCTTGGTCACCTTCAGCTTGG-3') and wild-type (5'-AGATTGGAAGGGTCGAGCTG-3') alleles.

**Statistics.** All comparative analyses were performed using unpaired Student's *t* tests. Pearson's chi-square test was performed using online software (<http://vassarstats.net/>).

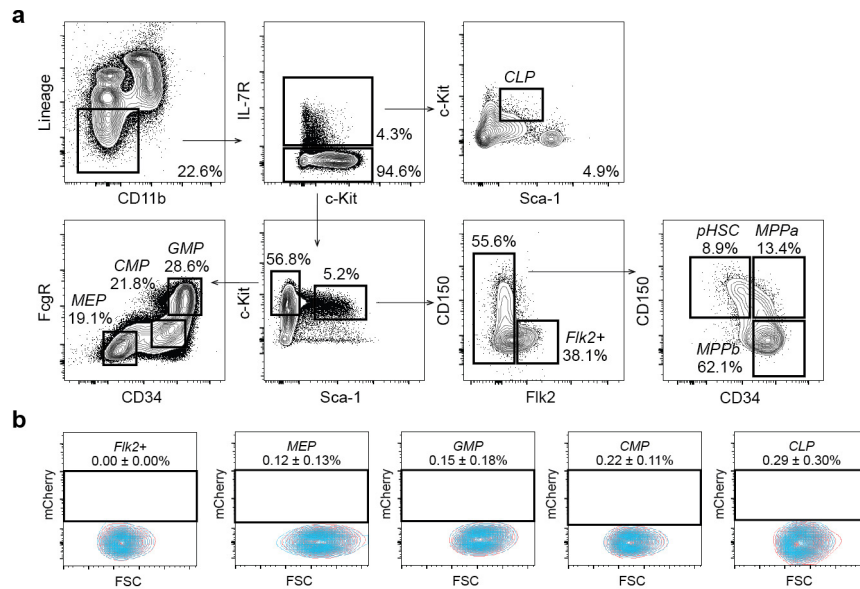
- Moraga, I. *et al.* Tuning cytokine receptor signaling by re-orienting dimer geometry with surrogate ligands. *Cell* **160**, 1196–1208 (2015).
- Wu, J., Anczuków, O., Krainer, A. R., Zhang, M. Q. & Zhang, C. Olego: fast and sensitive mapping of spliced mRNA-Seq reads using small seeds. *Nucleic Acids Res.* **41**, 5149–5163 (2013).
- Sanjana, N. E. *et al.* A transcription activator-like effector toolbox for genome engineering. *Nature Protocols* **7**, 171–192 (2012).
- Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).





**Extended Data Figure 1 | GEXC expression of previously reported HSC markers in mouse bone marrow. a,** Ideal expression pattern of HSC-specific genes (pink represents increased expression, blue represents decreased expression). **b,** Relative expression of *Hoxb5*

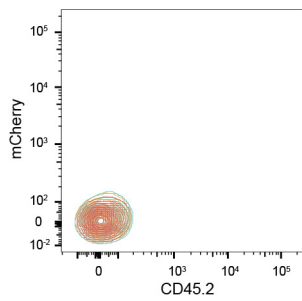
(top left), *α-catenin/Ctnn1* (top middle), *Fgd5* (top right), *CD150/Slamf1* (bottom left), *Hoxb4* (bottom middle), *Gfi-1* (bottom right) in haematopoietic and stromal populations as determined by microarray analysis.



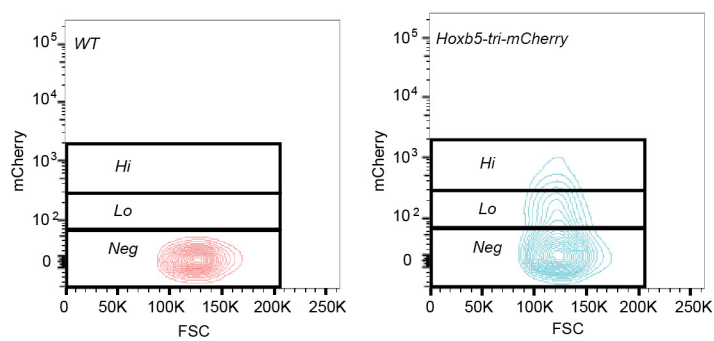
### Extended Data Figure 2 | Gating scheme for HSC and progenitors.

**a**, Representative flow cytometry gating to isolate pHSCs, MPPs, and oligopotent progenitors from mouse bone marrow. Panels gated as shown after exclusion of doublets and dead cells. **b**, *Hoxb5* reporter expression (red) in Flk2<sup>+</sup> MPPs, megakaryocyte erythrocyte progenitor (MEP),

granulocyte monocyte progenitor (GMP), common myeloid progenitor (CMP), and common lymphoid progenitor (CLP) populations compared to wild-type controls (blue). Values indicate the percentage of mCherry<sup>+</sup> cells ± s.d. in each fraction for  $n = 3$  mice.

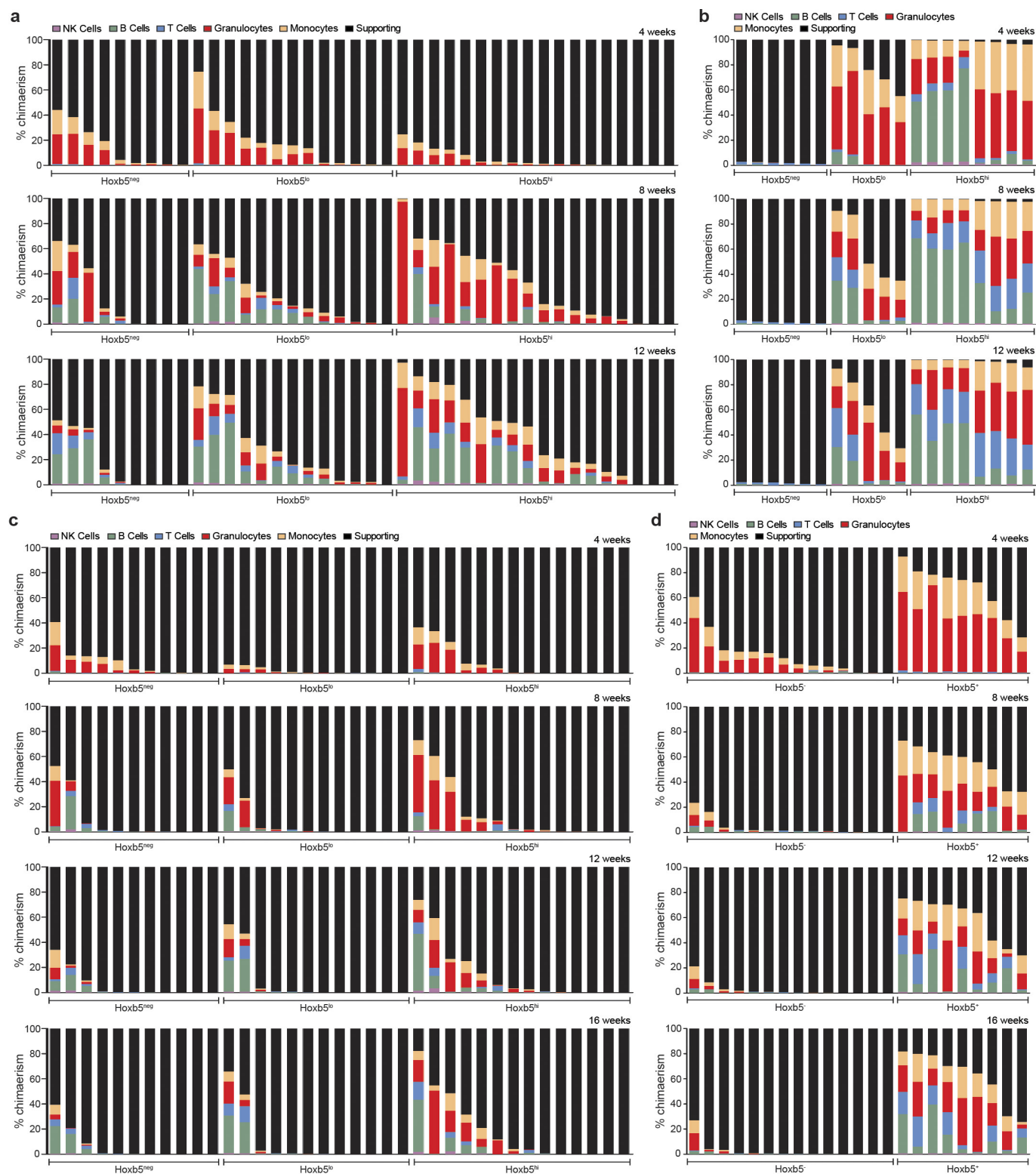


**Extended Data Figure 3 | *Hoxb5* is not expressed in CD45<sup>+</sup> bone marrow.** *Hoxb5* reporter expression in the CD45<sup>+</sup> compartment within bone marrow of wild-type (red) and three *Hoxb5*-tri-mCherry mice (blue, orange, and green,  $n = 3$  mice).



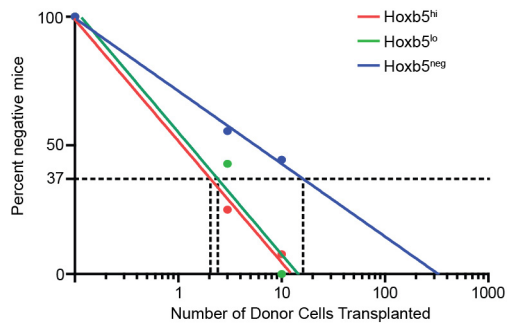
**Extended Data Figure 4 | FMO gating for  $Hoxb5^+$  signal.** Representative flow cytometry gating to separate mCherry ( $Hoxb5$ )-high, -low, and -negative populations in both wild-type and  $Hoxb5$ -tri-mCherry mice.



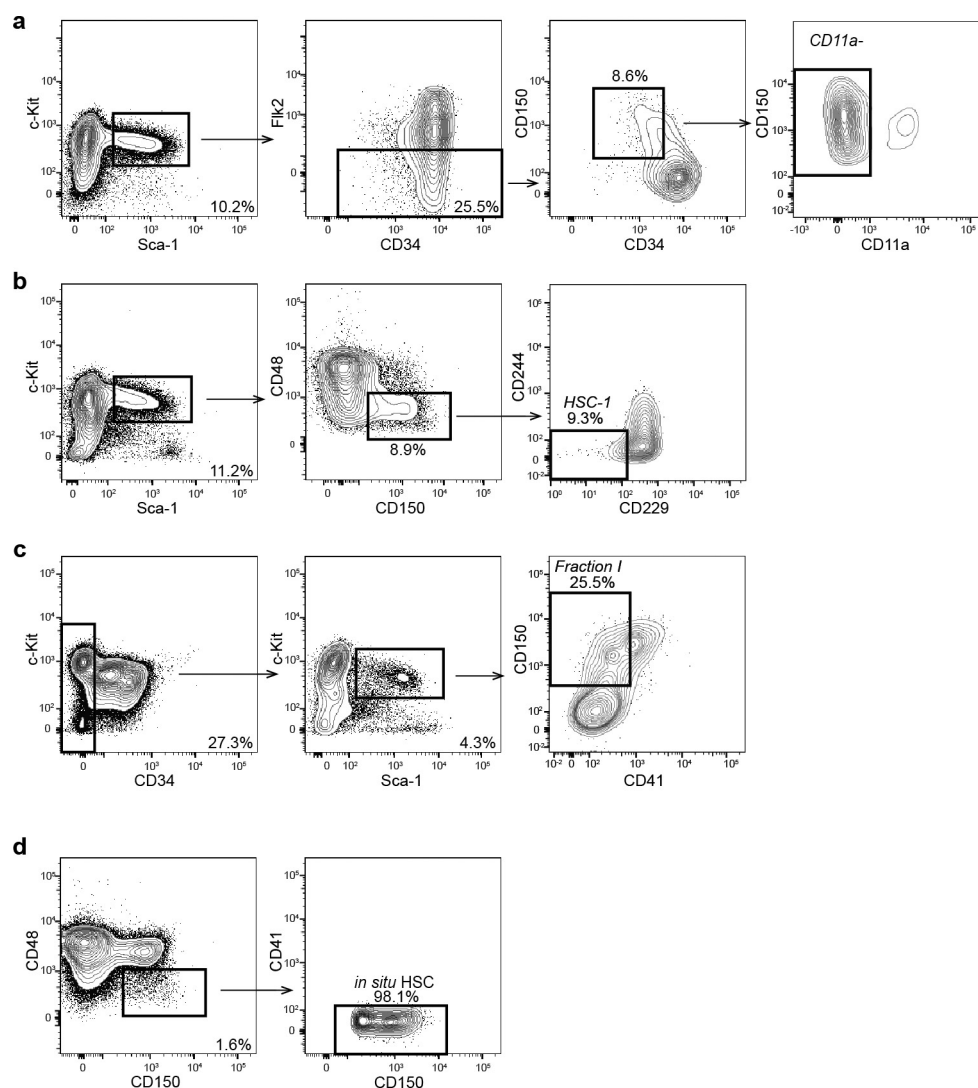


**Extended Data Figure 5 | Hoxb5 distinguishes between LT-HSCs and non-LT-HSCs. a,** Reconstitution kinetics in primary recipients 4, 8, and 12 weeks after receiving ten  $Hoxb5^{neg}$  ( $n = 9$  mice),  $Hoxb5^{lo}$  ( $n = 13$  mice), or  $Hoxb5^{hi}$  ( $n = 18$  mice) pHSCs. Each column represents an individual mouse. **b,** Reconstitution kinetics 4, 8, and 12 weeks after whole bone marrow secondary transplant. **c,** Reconstitution kinetics in primary

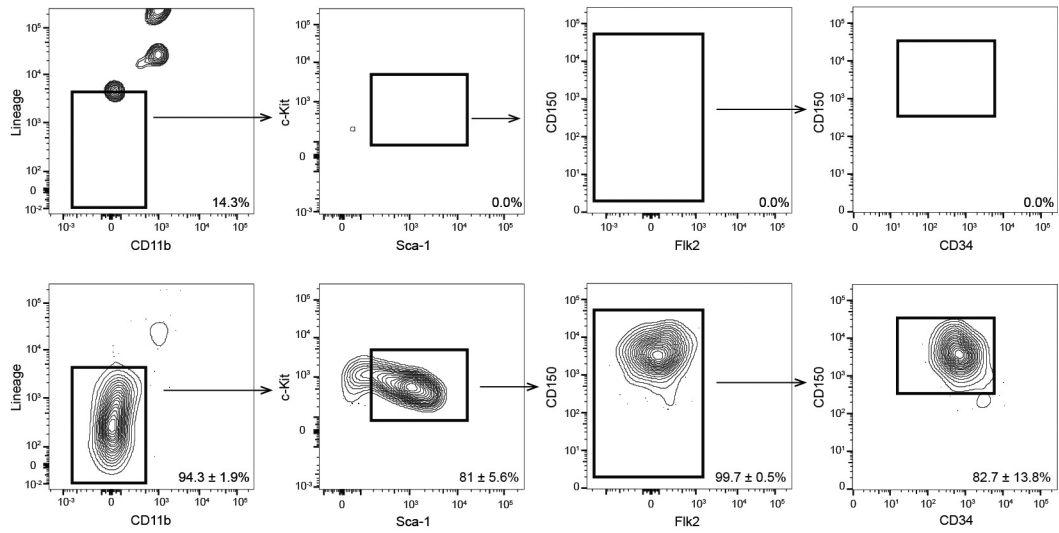
recipients receiving three  $Hoxb5^{neg}$  ( $n = 11$  mice),  $Hoxb5^{lo}$  ( $n = 12$  mice), or  $Hoxb5^{hi}$  ( $n = 14$  mice) pHSCs. Each column represents an individual mouse. **d,** Reconstitution kinetics following secondary transplant of 100 sorted LSK  $Hoxb5^{-}$  ( $n = 14$  mice) or  $Hoxb5^{+}$  ( $n = 9$  mice) cells and  $2 \times 10^5$  supporting cells.



**Extended Data Figure 6 | Limiting dilution analysis of Hoxb5<sup>+</sup> and Hoxb5<sup>-</sup> pHSCs.** Limiting dilution results of ten- and three-cell transplants of Hoxb5<sup>hi</sup> (red,  $n = 18$  mice for ten-cell and  $n = 14$  mice for three-cell), Hoxb5<sup>lo</sup> (green,  $n = 13$  mice for ten-cell and  $n = 12$  mice for three-cell), and Hoxb5<sup>neg</sup> (blue,  $n = 9$  mice for ten-cell and  $n = 11$  mice for three-cell). Frequency of LT/ST-HSCs by limiting dilution analysis is 1 in 2.1 for Hoxb5<sup>hi</sup>, 1 in 2.4 for Hoxb5<sup>lo</sup>, and 1 in 16.1 for Hoxb5<sup>neg</sup> cells.

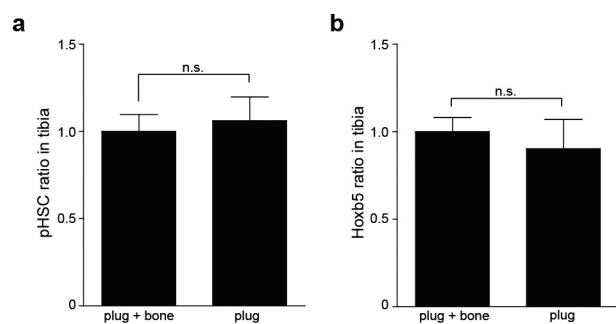


**Extended Data Figure 7 | Previously defined HSC immunophenotypes contain *Hoxb5*<sup>-</sup> cells.** Representative HSC gating strategy for various HSC definitions after exclusion of doublets and dead cells. **a**, CD11a<sup>-</sup> (LSK CD150<sup>+</sup>CD34<sup>-/lo</sup>CD11a<sup>-</sup>)<sup>21</sup>. **b**, HSC-1 (LSK CD150<sup>+</sup>CD48<sup>-/lo</sup>CD229<sup>-/lo</sup>CD244<sup>-</sup>)<sup>20</sup>. **c**, Fraction I (LSK CD150<sup>+</sup>CD34<sup>-/lo</sup>CD41<sup>-</sup>)<sup>18</sup>. **d**, CD150<sup>+</sup>CD48<sup>-</sup>CD41<sup>-</sup> cells<sup>22</sup> ( $n = 5$  mice).

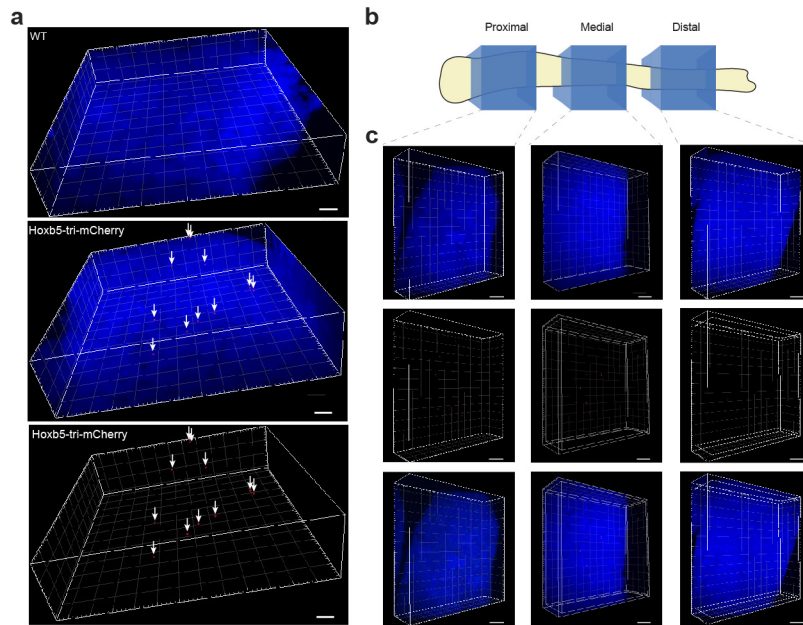


**Extended Data Figure 8 | Specificity of *Hoxb5* as a single marker for LT-HSCs. a,** Flow cytometry plots of wild type (top row) and *Hoxb5*-tri-mCherry (bottom row) after excluding doublets, dead cells, autofluorescence, and gating on *Hoxb5*<sup>+</sup> events. Frequencies shown are percentage in gate ± s.d. in each fraction ( $n = 3$  mice).





**Extended Data Figure 9 | Comparison of processing methods on pHSC and Hoxb5<sup>+</sup> LT-HSC yield. a, b,** Relative frequency of pHSCs (a) and Hoxb5<sup>+</sup> LT-HSCs (b) in tibial plugs (flushed) ( $n = 6$  mice) compared to tibial plugs plus bones (crushed) ( $n = 6$  mice).



**Extended Data Figure 10 |  $Hoxb5^+$  HSCs are evenly distributed in the tibia.** **a**, Distribution of  $Hoxb5^+$  cells (red and arrows) in bone marrow in 3D-reconstructed images. Nuclei are counterstained with DAPI (blue) wild-type (top panel) *Hoxb5*-tri-mCherry (middle and bottom panel). Scale bar, 100  $\mu\text{m}$ . **b**, Cartoon representing the location of the proximal,

medial, and distal sampling. **c**, Representative 3D-reconstructed images of  $Hoxb5^+$  cells (red) in proximal (left column), medial (middle column), and distal (right column) regions of the tibia. Scale bar, 150  $\mu\text{m}$ . Nuclei are counterstained with DAPI (blue);  $n = 3$  mice.

# Real-time, portable genome sequencing for Ebola surveillance

Joshua Quick<sup>1\*</sup>, Nicholas J. Loman<sup>1\*</sup>, Sophie Durauffour<sup>2,3\*</sup>, Jared T. Simpson<sup>4,5\*</sup>, Ettore Severi<sup>6\*</sup>, Lauren Cowley<sup>7\*</sup>, Joseph Akoi Bore<sup>2</sup>, Raymond Koundouno<sup>2</sup>, Gytis Dudas<sup>8</sup>, Amy Mikhail<sup>7</sup>, Nobila Ouédraogo<sup>9</sup>, Babak Afrough<sup>2,10</sup>, Amadou Bah<sup>2,11</sup>, Jonathan H. J. Baum<sup>2,3</sup>, Beate Becker-Ziaja<sup>2,3</sup>, Jan Peter Boettcher<sup>2,12</sup>, Mar Cabeza-Cabrerizo<sup>2,3</sup>, Álvaro Camino-Sánchez<sup>2</sup>, Lisa L. Carter<sup>2,13</sup>, Juliane Doerrbecker<sup>2,3</sup>, Theresa Enkirch<sup>2,14</sup>, Isabel García-Dorival<sup>2,15</sup>, Nicole Hetzelt<sup>2,12</sup>, Julia Hinzmann<sup>2,12</sup>, Tobias Holm<sup>2,3</sup>, Liana Eleni Kafetzopoulou<sup>2,16</sup>, Michel Koropogui<sup>2,17</sup>, Abigael Kosgey<sup>2,18</sup>, Eeva Kuisma<sup>2,10</sup>, Christopher H. Logue<sup>2,10</sup>, Antonio Mazzarelli<sup>2,19</sup>, Sarah Meisel<sup>2,3</sup>, Marc Mertens<sup>2,20</sup>, Janine Michel<sup>2,12</sup>, Didier Ngabo<sup>2,10</sup>, Katja Nitzsche<sup>2,3</sup>, Elisa Pallasch<sup>2,3</sup>, Livia Victoria Patrono<sup>2,3</sup>, Jasmine Portmann<sup>2,21</sup>, Johanna Gabriella Repits<sup>2,22</sup>, Natasha Y. Rickett<sup>2,15,23</sup>, Andreas Sachse<sup>2,12</sup>, Katrin Singethan<sup>2,24</sup>, Inês Vitoriano<sup>2,10</sup>, Rahel L. Yemanaberhan<sup>2,3</sup>, Elsa G. Zekeng<sup>2,15,23</sup>, Trina Racine<sup>25</sup>, Alexander Bello<sup>25</sup>, Amadou Alpha Sall<sup>26</sup>, Ousmane Faye<sup>26</sup>, Oumar Faye<sup>26</sup>, N'Faly Magassouba<sup>27</sup>, Cecelia V. Williams<sup>28,29</sup>, Victoria Amburgey<sup>28,29</sup>, Linda Winona<sup>28,29</sup>, Emily Davis<sup>29,30</sup>, Jon Gerlach<sup>29,30</sup>, Frank Washington<sup>29,30</sup>, Vanessa Monteil<sup>31</sup>, Marine Jourdain<sup>31</sup>, Marion Berer<sup>31</sup>, Alimou Camara<sup>31</sup>, Hermann Somlare<sup>31</sup>, Abdoulaye Camara<sup>31</sup>, Marianne Gerard<sup>31</sup>, Guillaume Bado<sup>31</sup>, Bernard Baillet<sup>31</sup>, Déborah Delaune<sup>32,33</sup>, Koumpingnin Yacouba Nebie<sup>34</sup>, Abdoulaye Diarra<sup>34</sup>, Yacouba Savane<sup>34</sup>, Raymond Bernard Pallawo<sup>34</sup>, Giovanna Jaramillo Gutierrez<sup>35</sup>, Natacha Milhano<sup>6,36</sup>, Isabelle Roger<sup>34</sup>, Christopher J. Williams<sup>6,37</sup>, Facinet Yattara<sup>17</sup>, Kuiama Lewandowski<sup>10</sup>, James Taylor<sup>38</sup>, Phillip Rachwal<sup>38</sup>, Daniel J. Turner<sup>39</sup>, Georgios Pollakis<sup>15,23</sup>, Julian A. Hiscox<sup>15,23</sup>, David A. Matthews<sup>40</sup>, Matthew K. O'Shea<sup>41</sup>, Andrew McD. Johnston<sup>41</sup>, Duncan Wilson<sup>41</sup>, Emma Hutley<sup>42</sup>, Erasmus Smit<sup>43</sup>, Antonino Di Caro<sup>2,19</sup>, Roman Wölfel<sup>2,44</sup>, Kilian Stoeker<sup>2,44</sup>, Erna Fleischmann<sup>2,44</sup>, Martin Gabriel<sup>2,3</sup>, Simon A. Weller<sup>38</sup>, Lamine Koivogui<sup>45</sup>, Boubacar Diallo<sup>34</sup>, Sakoba Keita<sup>17</sup>, Andrew Rambaut<sup>8,46,47</sup>, Pierre Formenty<sup>34</sup>, Stephan Günther<sup>2,3</sup> & Miles W. Carroll<sup>2,10,48,49</sup>

**The Ebola virus disease epidemic in West Africa is the largest on record, responsible for over 28,599 cases and more than 11,299 deaths<sup>1</sup>. Genome sequencing in viral outbreaks is desirable to characterize the infectious agent and determine its evolutionary rate. Genome sequencing also allows the identification of signatures of host adaptation, identification and monitoring of diagnostic targets, and characterization of responses to vaccines and treatments. The Ebola virus (EBOV) genome substitution rate in the Makona strain has been estimated at between  $0.87 \times 10^{-3}$  and  $1.42 \times 10^{-3}$  mutations per site per year. This is equivalent to 16–27 mutations in each genome, meaning that sequences diverge rapidly enough to identify distinct sub-lineages during a prolonged epidemic<sup>2–7</sup>. Genome sequencing provides a high-resolution view of pathogen evolution and is increasingly sought after for outbreak surveillance. Sequence data may be used to guide control measures, but only if the results are generated quickly enough to inform interventions<sup>8</sup>. Genomic surveillance during the epidemic has been sporadic**

**owing to a lack of local sequencing capacity coupled with practical difficulties transporting samples to remote sequencing facilities<sup>9</sup>. To address this problem, here we devise a genomic surveillance system that utilizes a novel nanopore DNA sequencing instrument. In April 2015 this system was transported in standard airline luggage to Guinea and used for real-time genomic surveillance of the ongoing epidemic. We present sequence data and analysis of 142 EBOV samples collected during the period March to October 2015. We were able to generate results less than 24 h after receiving an Ebola-positive sample, with the sequencing process taking as little as 15–60 min. We show that real-time genomic surveillance is possible in resource-limited settings and can be established rapidly to monitor outbreaks.**

Conventional sequencing technologies are difficult to deploy in developing countries, where availability of continuous power and cold chains, laboratory space, and trained personnel is restricted. In addition, some genome sequencer instruments, such as those using optical

<sup>1</sup>Institute of Microbiology and Infection, University of Birmingham, Birmingham B15 2TT, UK. <sup>2</sup>The European Mobile Laboratory Consortium, Bernhard-Nocht-Institute for Tropical Medicine, D-20359 Hamburg, Germany. <sup>3</sup>Bernhard-Nocht-Institute for Tropical Medicine, D-20359 Hamburg, Germany. <sup>4</sup>Ontario Institute for Cancer Research, Toronto M5G 0A3, Canada. <sup>5</sup>Department of Computer Science, University of Toronto, Toronto M5S 3G4, Canada. <sup>6</sup>European Centre for Disease Prevention and Control (ECDC), 171 65 Solna, Sweden. <sup>7</sup>National Infection Service, Public Health England, London NW9 5EQ, UK. <sup>8</sup>Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 2FL, UK. <sup>9</sup>Postgraduate Training for Applied Epidemiology (PAE, German FETP), Robert Koch Institute, D-13302 Berlin, Germany. <sup>10</sup>National Infection Service, Public Health England, Porton Down, Wiltshire SP4 0JG, UK. <sup>11</sup>Swiss Tropical and Public Health Institute, 4002 Basel, Switzerland. <sup>12</sup>Robert Koch Institute, D-13302 Berlin, Germany. <sup>13</sup>University College London, London WC1E 6BT, UK. <sup>14</sup>Paul-Ehrlich-Institut, Division of Veterinary Medicine, D-63225 Langen, Germany. <sup>15</sup>Institute of Infection and Global Health, University of Liverpool, Liverpool L69 7BE, UK. <sup>16</sup>Laboratory for Clinical and Epidemiological Virology, Department of Microbiology and Immunology, KU Leuven, Leuven B-3000, Belgium. <sup>17</sup>Ministry of Health Guinea, Conakry BP 585, Guinea. <sup>18</sup>Kenya Medical Research Institute, Nairobi P.O. BOX 54840 - 00200, Kenya. <sup>19</sup>National Institute for Infectious Diseases L. Spallanzani, 00149 Rome, Italy. <sup>20</sup>Friedrich-Loeffler-Institute, D-17493 Greifswald, Germany. <sup>21</sup>Federal Office for Civil Protection, Spiez Laboratory, 3700 Spiez, Switzerland. <sup>22</sup>Janssen-Cilag, Stockholm, Box 7073 – 19207, Sweden. <sup>23</sup>NIHR Health Protection Research Unit in Emerging and Zoonotic Infections, University of Liverpool, Liverpool L69 7BE, UK. <sup>24</sup>Institute of Virology, Technische Universität München, D-81675 Munich, Germany. <sup>25</sup>Public Health Agency of Canada, Winnipeg, Manitoba R3E 3R2, Canada. <sup>26</sup>Institut Pasteur Dakar, Dakar, DP 220 Senegal. <sup>27</sup>Laboratoire de Fièvres Hémorragiques de Guinée, Conakry BP 5680, Guinea. <sup>28</sup>Sandia National Laboratories, PO Box 5800 MS1363, Albuquerque, New Mexico 87185-1363, USA. <sup>29</sup>Ratoma Ebola Diagnostic Center, Conakry, Guinea. <sup>30</sup>MRI Global, Kansas City, Missouri 64110-2241, USA. <sup>31</sup>Expertise France, Laboratoire K-plan de Forcariah en Guinée, 75006 Paris, France. <sup>32</sup>Fédération des Laboratoires - HIA Bégin, 94163 Saint-Mandé cedex, France. <sup>33</sup>Laboratoire de Biologie - Centre de Traitement des Soignants, Conakry, Guinea. <sup>34</sup>World Health Organization, Conakry BP 817, Guinea. <sup>35</sup>London School of Hygiene and Tropical Medicine, London EC1E 7HT, UK. <sup>36</sup>Norwegian Institute of Public Health, PO Box 4404 Nydalen, 0403 Oslo, Norway. <sup>37</sup>Public Health Wales, Cardiff CF11 9LJ, UK. <sup>38</sup>Defence Science and Technology Laboratory (Dstl) Porton Down, Salisbury SP4 0JQ, UK. <sup>39</sup>Oxford Nanopore Technologies, Oxford OX4 4GA, UK. <sup>40</sup>Department of Cellular and Molecular Medicine, School of Medical Sciences, University of Bristol, Bristol BS8 1TD, UK. <sup>41</sup>Academic Department of Military Medicine, Royal Centre for Defence Medicine, Birmingham B15 2TH, UK. <sup>42</sup>Centre of Defence Pathology, Royal Centre for Defence Medicine, Birmingham B15 2TH, UK. <sup>43</sup>Queen Elizabeth Hospital, Birmingham B15 2TH, UK. <sup>44</sup>Bundeswehr Institute of Microbiology, D-80937 Munich, Germany. <sup>45</sup>Institut National de Santé Publique, Conakry BP 1147, Guinea. <sup>46</sup>Fogarty International Center, National Institutes of Health, Bethesda, MD 20892-2220, USA. <sup>47</sup>Centre for Immunology, Infection and Evolution, University of Edinburgh, Edinburgh EH9 2FL, UK. <sup>48</sup>University of Southampton, South General Hospital, Southampton SO16 6YD, UK. <sup>49</sup>NIHR Health Protection Research Unit in Emerging and Zoonotic Infections, PHE Porton Down, UK.

\*These authors contributed equally to this work.

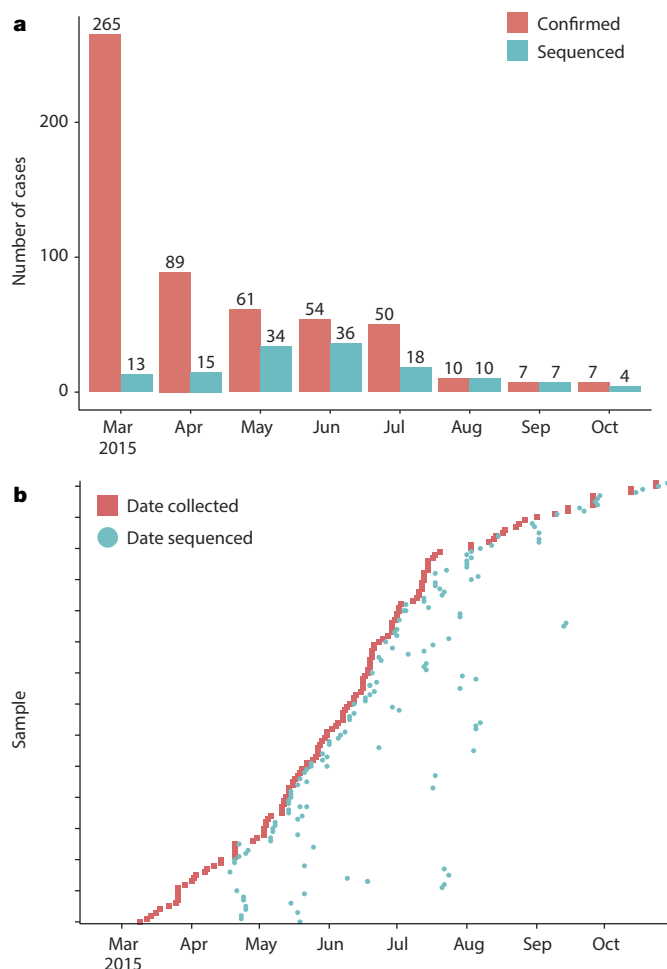


**Figure 1 | Deployment of the portable genome surveillance system in Guinea.** **a**, We were able to pack all instruments, reagents and disposable consumables within aircraft baggage. **b**, We initially established the genomic surveillance laboratory in Donka Hospital, Conakry, Guinea. **c**, Later we moved the laboratory to a dedicated sequencing laboratory in Coyah prefecture. **d**, Within this laboratory we separated the sequencing instruments (on the left) from the PCR bench (to the right). An uninterruptable power supply can be seen in the middle that provides power to the thermocycler. (Photographs taken by J.Q. and S.D.)

readings, for example the Illumina platform, require precise microscope alignment and repeated calibration by trained engineers<sup>7,10</sup>. Recently, a new highly portable genome sequencer has become available. The MinION (Oxford Nanopore Technologies, Oxford, UK) weighs less than 100g. Data are read off the MinION from a laptop via a Universal Serial Bus (USB) port from which the instrument also draws power. The MinION works by taking frequent electrical current measurements as a single strand of DNA passes through a protein nanopore at 30 bases per second. DNA strands in the pore disrupts ionic flow, resulting in detectable changes in current that is dependent on the nucleotide sequence. Because the MinION detects single molecules it has a much higher error rate (between 10–20%<sup>11,12</sup>) than high-throughput instruments that read clonal copies of DNA molecules. Single-molecule sequencing has the advantage of being able to read extremely long molecules of DNA (50 kb or longer<sup>12,13</sup>). In order to generate accurate sequences, genomic regions must be read many times, with errors eliminated through consensus averaging. This system has previously been used to investigate a bacterial outbreak, but not yet a viral outbreak<sup>14</sup>.

We designed a laboratory protocol to permit EBOV genome sequencing on the MinION that employed a targeted reverse transcriptase polymerase chain reaction (RT-PCR) in order to isolate sufficient DNA for sequencing. We considered and rejected an alternative approach, that of total RNA sequencing, as this approach also amplifies human-derived transcripts and dilutes viral signal<sup>15</sup>. We designed a panel of 38 primer pairs that would span the EBOV genome (Extended Data Fig. 1a, Supplementary Table 1). In pilot experiments at the Defence Science and Technology Laboratory (Dstl) Porton Down, UK, we sequenced a historic Zaire Ebolavirus using MinION as well as the Illumina MiSeq. Due to difficulties obtaining equal balancing of each of the 38 amplicon pairs only 65.7% of the EBOV genome was covered by at least 25 reads, compared with 87.4% on Illumina. However, nucleotide variants in those highly covered regions were concordant with those obtained from Illumina sequencing, with the exception of a single variant in a homopolymeric region. MinION sequencing currently cannot easily resolve the length of homopolymers of 5 bases or greater<sup>16</sup>.

Next we designed a genome surveillance system that could be transported to West Africa. The system consisted of three MinION instruments (Oxford Nanopore Technologies, UK), four laptops,



**Figure 2 | Real-time genomics surveillance in context of the Guinea Ebola virus disease epidemic.** **a**, Here we show the number of reported cases of Ebola virus disease in Guinea (red) in relation to the number of EBOV new patient samples ( $n = 137$ , in blue) generated during this study. **b**, For each of the 142 sequenced samples, we show the relationship between sample collection date (red) and the date of sequencing (blue). Twenty-eight samples were sequenced within three days of the sample being taken, and sixty-eight samples within a week. Larger gaps represent retrospective sequencing of cases to provide additional epidemiological context.

a thermocycler, a heat block, pipettes and sufficient reagents and consumables to perform sequencing (a full list of equipment is shown in Extended Data Fig. 2). We were able to pack this into less than 50 kg of standard airline travel luggage (Fig. 1a). We initially installed the genome surveillance system in the European Mobile Laboratory in Donka Hospital in Conakry, Guinea (Fig. 1b). Later on, the equipment was moved to a dedicated laboratory, located within the Coyah Ebola Treatment Unit (Fig. 1c, d).

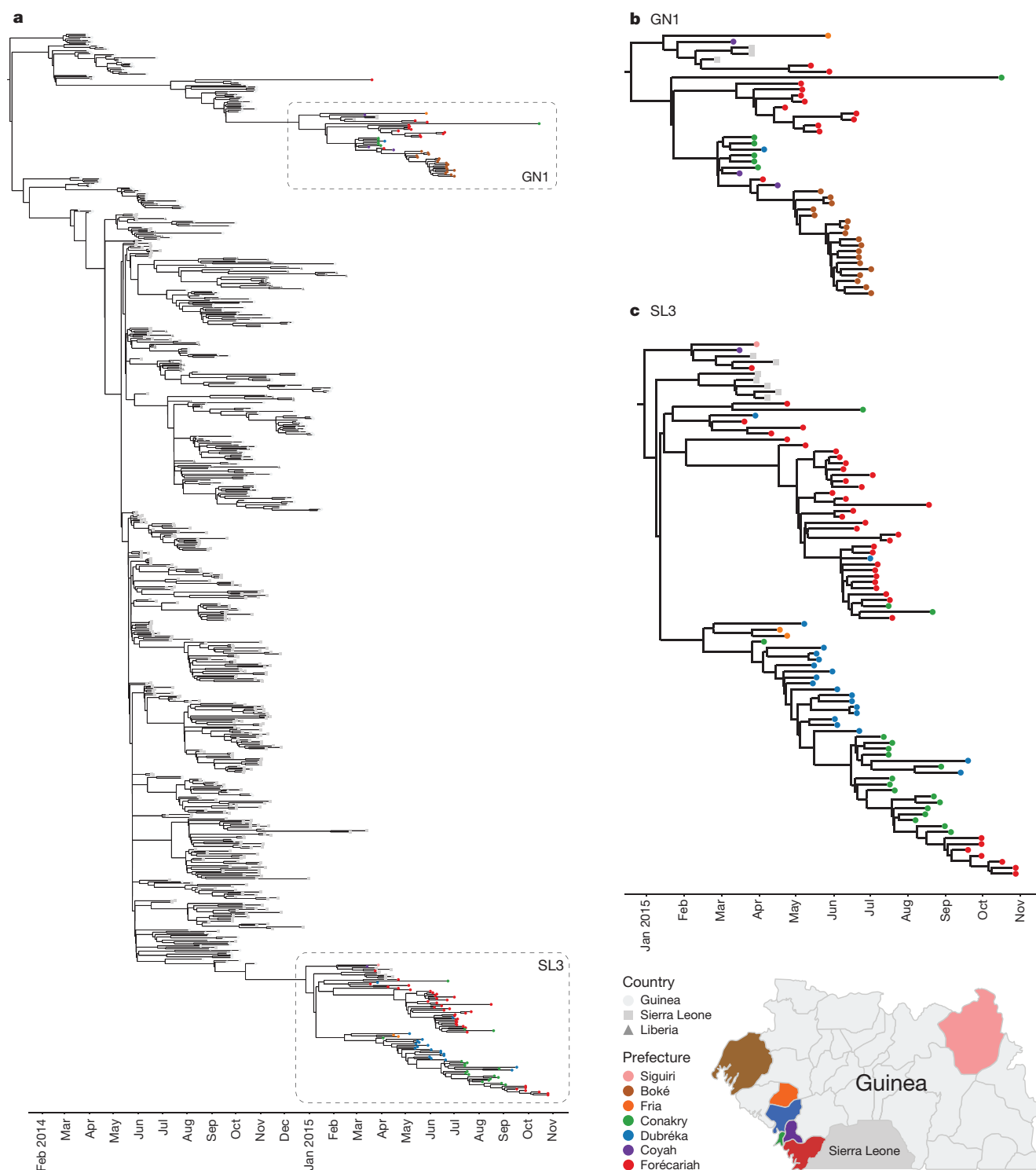
We started sequencing genomes within 2 days of arriving in Guinea. We found early on that we were able to reliably generate long amplicons (around 2 kb in length) using primer pairs (Supplementary Table 4) in different combinations (Extended Data Fig. 1b, c). Using as few amplicons as possible significantly reduces effort when preparing samples. We found a combination of 11 amplicons that reliably amplified >97% of the EBOV genome.

We developed a bioinformatics approach that would yield accurate genotypes, and validated this using Makona virus samples from a previous study<sup>3</sup>. The bioinformatics workflow is detailed in the Methods and summarized in Extended Data Fig. 3. This validation process demonstrated that our bioinformatics analysis approach was robust. We compared our consensus sequences to those generated using Illumina



sequencing and found that our approach was highly concordant, with no false positive variant calls. In several cases, we were unable to determine variants because they fell either within the primer binding region, or they were outside of the regions of the EBOV genome covered by our amplicon set (Extended Data Fig. 4a). These positions are represented as ambiguous nucleotides in the final consensus sequences used

for analysis. Despite these masked positions, phylogenetic inference showed that samples clustered identically (Extended Data Fig. 4b). We determined that, despite the instrument's high error rate, use of electrical current information meant that 25-fold read coverage of genome positions was sufficient to determine accurate genotypes (Extended Data Fig. 5).



**Figure 3 | Evolution of EBOV over the course of the Ebola virus disease epidemic.** **a**, Time-scaled phylogeny of 603 published sequences with 125 high quality sequences from this study. The shape of nodes on the tree demonstrates country of origin. Our results show Guinean samples (coloured circles) belong to two previously identified lineages, GN1 and SL3. **b**, GN1 is deeply branching with early epidemic samples. **c**, SL3 is

related to cases identified in Sierra Leone. Samples are frequently clustered by geography (indicated by colour of circle) and this provides information as to origins of new introductions, such as in the Boké epidemic in May 2015. Map figure adapted from SimpleMaps website (<http://simplemaps.com/resources/svg-gn>).

After deployment of the genome surveillance system, we worked in partnership with diagnostic laboratories in Guinea to provide real-time sequencing results to National Coordination in Guinea and the World Health Organisation (WHO). Collaborating laboratories provided left-over diagnostic RNA extracts for sequencing. The genome sequencing workflow, including amplification, sequence library preparation and sequencing could be accomplished within a working day. In one case, including remote bioinformatics analysis, the fastest time from patient sample to answer was achieved in less than 24 h (Supplementary Table 1), although the protocol was more usually performed over two working days. We found that in half of cases, we were able to generate sufficient reads on the MinION (between around 5,000 and 10,000) in less than an hour (Extended Data Fig. 6). In total, 142 samples were sequenced over 148 MinION runs during the 6-month period, providing extensive coverage of reported cases in the outbreak (Fig. 2). Full details of samples and runs are in the Supplementary Data. We failed to generate amplicons for some samples, resulting in missing regions of the genome. Such samples often corresponded to those with a high RT-PCR cycle threshold ( $C_t$ ) value, suggestive of lower viral loads (Extended Data Fig. 7). For these we used a modified RT-PCR scheme using 19 shorter amplicons. We assumed that difficulties generating long amplicons related to low numbers of starting molecules of that length in the original sample. We excluded 17 samples owing to quality control issues, for example single nucleotide polymorphism (SNP) calling sensitivity of less than 75%. We found that in-field performance of the system was comparable with validation experiments performed in the UK, suggesting that the system tolerated transportation well (Extended Data Fig. 8).

We combined our sequencing data set with 603 samples from other studies and inferred a time-scaled phylogenetic tree using the BEAST software package (Fig. 3). A maximum likelihood analysis and root-to-tip analysis showed good agreement between sampling date and root-to-tip divergence (Extended Data Figs 9 and 10a). We estimated a substitution rate of  $1.19 \times 10^{-3}$  (95% interval,  $1.09 \times 10^{-3}$ ,  $1.29 \times 10^{-3}$ ) of the combined data set (Extended Data Fig. 10b). This is consistent with rates from previous studies<sup>2–7</sup>. Results generated within the first 10 days of starting real-time sequencing indicated that the persisting Guinean cases belonged to two major lineages, named GN1 and SL3, that had been established near the beginning of the epidemic (Fig. 3). Lineage GN1 is deeply branching from early cases in Guinea and has been infrequently seen in Sierra Leone<sup>2</sup>, suggesting that it has been largely confined to Guinea. The second lineage identified here was derived from lineage SL3 which was first detected in Sierra Leone by Gire *et al.*<sup>2</sup>, but was later seen circulating in Conakry towards the end of 2014<sup>3</sup>. Through integration of our data set with those generated by a different group operating in Sierra Leone we detected that both GN1 and SL3 had also been seen in Sierra Leone early in 2015, suggestive of transmission between the countries<sup>17</sup>.

This work demonstrates a step change in our ability to perform genomic surveillance prospectively during outbreaks under resource-limited conditions. However, numerous obstacles remain before such genomically informed investigations are routine. In practical terms, we encountered significant logistical issues when performing this work, notably the absence of reliable, continuous mains electrical power, forcing a dependence on unreliable electrical generators and uninterruptable power supply (UPS) units, particularly for the bulky PCR thermocyclers. However, portable, battery-powered thermocyclers are in development, and isothermal approaches may be preferable for future work<sup>18</sup>. By contrast, the MinION sequencer was unaffected by power outages and surges. We faced consistent issues with internet connectivity, which is currently required for analysis. There is a pressing need for a fully offline version of the analysis presented here. This would reduce the dependence on high bandwidth connections. However it is likely that phylogenetic analysis will continue to be performed remotely (discussed further in the supplementary Field Guide to Portable Sequencing). In this analysis we focused on variant calling

approaches. A *de novo* approach to analysis would be preferable, but this would currently result in insertion and deletion errors due to poor resolution of homopolymeric tracts on the MinION. Our approach relies on amplification of genetic material before sequencing. In other epidemics, where the causative pathogen may be unidentified this is a drawback due to the need to have a priori knowledge of the pathogen genome sequence. In this event, sequencing directly from clinical material may be better, although sensitivity issues persist<sup>15</sup>.

Real-time genomic surveillance is a new tool in our arsenal to assist difficult epidemiological investigations, and to provide an international and environmental context to emerging infectious diseases. This may improve the efficiency of resource allocation and the timeliness of epidemiological investigations through genomically informed investigations of transmission chains. Real-time genomic surveillance also increases the possibility of identifying previously unknown chains of transmission. By integrating in real time our data set with that of a second group performing sequencing in Sierra Leone, we identified evidence of frequent transmissions across the border with Guinea. Crucially, we released data at regular intervals throughout this project through Github, integrating our results with those of others, displayed interactively at <http://ebola.nextstrain.org>. We employed the Virological web forum to discuss complex cases (<http://virological.org>). This system will continue to support the West African epidemic response and will serve as a template for genomic surveillance of future outbreaks.

The Ebola epidemic was officially declared to be over on 14 January 2016 (<http://www.who.int/mediacentre/news/releases/2016/ebola-zero-liberia/en/>). Hours later, a new case of EVD was confirmed in Sierra Leone (<http://who.int/mediacentre/news/statements/2016/new-ebola-case/en/>), confirming warnings that further flare-ups may be expected. Such cases pose pressing questions about their source that may be answered through genomic surveillance, by determining links to previously infected individuals<sup>19</sup> and ruling out a new zoonotic spillover event. We now stand poised to answer such questions quickly.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 18 November 2015; accepted 15 January 2016.**

**Published online 3 February 2016.**

- World Health Organisation. Ebola Situation Report - 11 November 2015. <http://apps.who.int/ebola/current-situation/ebola-situation-report-11-november-2015> (2015).
- Gire, S. K. *et al.* Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* **345**, 1369–1372 (2014).
- Carroll, M. W. *et al.* Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa. *Nature* **524**, 97–101 (2015).
- Simon-Loriere, E. *et al.* Distinct lineages of Ebola virus in Guinea during the 2014 West African epidemic. *Nature* **524**, 102–104 (2015).
- Park, D. J. *et al.* Ebola virus epidemiology, transmission, and evolution during seven months in Sierra Leone. *Cell* **161**, 1516–1526 (2015).
- Tong, Y.-G. *et al.* Genetic diversity and evolutionary dynamics of Ebola virus in Sierra Leone. *Nature* **524**, 93–96 (2015).
- Kugelman, J. R. *et al.* Monitoring of Ebola Virus Makona evolution through establishment of advanced genomic capability in Liberia. *Emerg. Infect. Dis.* **21**, 1135–1143 (2015).
- Gardy, J., Loman, N. J. & Rambaut, A. Real-time digital pathogen surveillance—the time is now. *Genome Biol.* **16**, 155 (2015).
- Yozwiak, N. L., Schaffner, S. F. & Sabeti, P. C. Data sharing: make outbreak research open access. *Nature* **518**, 477–479 (2015).
- Heger, M. Liberia's LIBR Genome Center monitors Ebola outbreak, emerging pathogens. <https://www.genomeweb.com/sequencing-technology/liberias-libr-genome-center-monitors-ebola-outbreak-emerging-pathogens> (2015).
- Quick, J., Quinlan, A. R. & Loman, N. J. A reference bacterial genome dataset generated on the MinION™ portable single-molecule nanopore sequencer. *Gigascience* **3**, 22 (2014).
- Jain, M. *et al.* Improved data analysis for the MinION nanopore sequencer. *Nature Methods* **12**, 351–356 (2015).
- Urban, J. M., Bliss, J., Lawrence, C. E. & Gerbi, S. A. Sequencing ultra-long DNA molecules with the Oxford Nanopore MinION. Preprint at <http://biorxiv.org/content/early/2015/05/13/019281> (2015).
- Quick, J. *et al.* Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. *Genome Biol.* **16**, 114 (2015).

15. Greninger, A. L. *et al.* Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome Med.* **7**, 99 (2015).
16. Loman, N. J., Quick, J. & Simpson, J. T. A complete bacterial genome assembled *de novo* using only nanopore sequencing data. *Nature Methods* **12**, 733–735 (2015).
17. Goodfellow, I. *et al.* Recent evolution patterns of Ebola virus obtained by direct sequencing in Sierra Leone. <http://virological.org/t/recent-evolution-patterns-of-ebola-virus-obtained-by-direct-sequencing-in-sierra-leone/150> (2015).
18. Herold, K. E., Sergeev, N., Matviyenko, A. & Rasooly, A. in *Biosensors and Biodetection* **504**, 441–458 (Humana, 2009).
19. Mate, S. E. *et al.* Molecular evidence of sexual transmission of Ebola virus. *N. Engl. J. Med.* **373**, 2448–2454 (2015).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** The EMLab is a technical partner in the WHO Emerging and Dangerous Pathogens Laboratory Network (EDPLN), and the Global Outbreak Alert and Response Network (GOARN) and the deployments in West Africa have been coordinated and supported by the GOARN Operational Support Team at WHO/HQ and the African Union. This work was carried out in the context of the project EVIDENT (Ebola virus disease: correlates of protection, determinants of outcome, and clinical management) that received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 666100 and in the context of service contract IFS/2011/272-372 funded by Directorate-General for International Cooperation and Development. J.Q. is funded by the NIHR Surgical Reconstruction and Microbiology Research Centre (SRMRC). N.J.L. is funded by a Medical Research Council Special Training Fellowship in Biomedical Informatics (to September 2015) and a Medical Research Council Bioinformatics Fellowship. J.T.S. is supported by the Ontario Institute for Cancer Research through funding provided by the Government of Ontario. Dstl support was funded by the UK Ministry of Defence (MOD). Dstl authors thank S. Lonsdale, C. Lonsdale and C. Mayers for supply of RNA, previous assistance, and review of the manuscript. The views expressed in this paper are not necessarily endorsed by the UK MOD. A.R. was supported by EU Seventh Framework Programme [FP7/2007-2013] under Grant Agreement no. 278433-PREDEMICS and ERC Grant agreement no. 260864. We are grateful for the generous support of University of

Birmingham alumni for donations in support of the pilot work. The MRC Cloud Infrastructure for Microbial Bioinformatics (CLIMB) cyberinfrastructure was used to conduct bioinformatics analysis. The authors would like to thank B. Oppenheim and C. Wardius for help with logistics and the staff of Alta Biosciences, University of Birmingham and Sigma-Aldrich for generating PCR primers especially rapidly for this project. The authors would like to thank scientists deployed from the Special Pathogens Program from the National Microbiology Laboratory, Public Health Agency of Canada, who worked on EBOV diagnostics in Guinea. We are grateful to I. Goodfellow, M. Cotten and P. Kellam for permission to include sequences from Sierra Leone in this analysis. We thank R. Vipond for assistance with validation experiments. We thank H. Eno and B. Myers for help with proofreading. We are thankful for the generous support of reagents and technical support from Oxford Nanopore. We thank the staff at Oxford Nanopore for technical and logistical support during this project with special thanks to S. Brooking, O. Hartwell, R. Pettett, C. Brown, G. Sanghera and R. Ronan. We thank T. Bedford and R. Neher for developing the Nextstrain website.

**Author Contributions** N.J.L., J.Q., M.K.O'S., D.W., S.G., M.W.C. conceived the study. N.J.L., J.Q., M.K.O'S., S.A.W., J.T., P.R., D.T. designed the lab in a suitcase and laboratory protocol and initial validation. J.Q., S.D., L.C., J.A.B., R.K., L.E.K., and A.Ma. performed MinION sequencing. N.J.L., J.Q. and J.T.S. performed bioinformatics analysis and wrote software. J.T.S. added variant calling support to the nanopolish software. N.J.L., J.Q., S.D., E.S., P.F., L.C., A.Mi., N.M. and I.R. analysed the data. G.D., A.R., N.J.L., J.Q. and G.P. performed phylogenetic analysis. J.A.H., D.A.M., G.P., K.L., B.A. assisted further validation experiments. M.W.C., M.Ga., S.G., A.D.C., K.S., E.F. and R.W. coordinated activities for the European Mobile Laboratories. N.J.L., J.Q., S.D., M.W.C., S.G., M.K.O'S., A.R., E.S., P.F., I.R., A.Mi., and L.C. wrote the manuscript. All other authors were involved either in sample collection, and/or logistical support and strategic oversight for the work.

**Author Information** MinION and Illumina raw sequence files have been deposited into the European Nucleotide Archive under project code PRJEB10571. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.J.L. ([n.j.loman@bham.ac.uk](mailto:n.j.loman@bham.ac.uk)).

## METHODS

**Ethics statement.** The National Committee of Ethics in Medical Research of Guinea (permit no. 11/CNERS/14) approved the use of diagnostic leftover samples and corresponding patient data for this study. As the samples had been collected as part of the public health response to control the outbreak in West Africa, informed consent was not obtained from patients.

**Transportation.** All equipment was loaded into a Pelican 1610 case (Pelican, Torrance, USA), cold chain reagents were packed into two polystyrene boxes with either ice or cool packs. These were sealed and placed in a holdall with the plastic consumables. Both pieces of luggage were flown by air as normal checked baggage.

**RNA extraction.** RNA was extracted from 50 µl whole blood, 140 µl serum, 140 µl of resuspended swab or 140 µl urine using the QIAamp Viral RNA Mini Kit (Qiagen, Manchester, UK), following the manufacturer's instructions. Samples were inactivated by adding 560 µl of Buffer AVL (Qiagen) and 560 µl of 100% ethanol while still in a glove box, this method has been shown to inactivate EBOV in blood samples<sup>20</sup>. Following inactivation, samples were handled on the bench employing standard laboratory safety precautions.

**RT-PCR.** Individual 25 µl RT-PCR reactions were performed using the SuperScript III One-Step RT-PCR System with Platinum Taq DNA Polymerase (Life Technologies Ltd, Paisley, UK). Each reaction was made up by adding 12.5 µl 2 × reaction mix, 1 µl enzyme mix, 1 µl primers (10 µM), 0.5 µl RNA extract and nuclease-free water. Thermocycling was performed on an Eppendorf Master Cycler Personal instrument with the following program: 60 °C for 30 min, 94 °C for 2 min followed by 45 cycles of 94 °C for 15 s, 55 °C for 30 s, 68 °C for 2 min and a final extension of 68 °C for 5 min.

**MinION library preparation.** Each reaction was quantified on a Qubit 3.0 fluorimeter using the dsDNA HS assay (Life Technologies). Equimolar amounts of each amplicon product to a total DNA mass of 1 µg was pooled into a single tube and cleaned-up using an equal volume of MAGBIO HighPrep PCR beads (AutoQ Biosciences, Reading, UK). Pooled amplicons were diluted to 85 µl, and end-repaired in a total volume of 100 µl, using the NEBNext End Repair Module (New England Biolabs, Hitchin, UK) before being cleaned up using an equal volume of HighPrep PCR beads and eluting in 25 µl nuclease-free water. 3' dA-tailing was performed using the NEBNext dA-Tailing Module (New England Biolabs) in a volume of 30 µl, before being cleaned up using an equal volume of HighPrep PCR beads and eluting in 30 µl nuclease-free water. 10 µl of 'Adaptor mix' and 10 µl 'HP adaptor' supplied in the SQK-MAP005 library preparation kit (Oxford Nanopore Technologies, Oxford, UK) were added to the dA-tailed amplicons along with 50 µl, Blunt/TA Ligase Master Mix (New England Biolabs) in a Protein LoBind tube (Eppendorf UK) and incubated for 10 min. The resulting sequencing library was purified using Dynabeads His-Tag Isolation and Pulldown beads (Life Technologies, Stevenage, UK) according to the SQK-MAP005 protocol supplied by Oxford Nanopore Technologies as part of the MinION Access Program. The final library was quantified using the Qubit to confirm the process had been successful. 6 µl of library was diluted using 75 µl '2x Running Buffer', 66 µl Nuclease-free water (Promega UK, Chilworth, UK) and 3 µl, 'Fuel Mix'.

**MinION sequencing.** A new flowcell was unpackaged and fitted onto the MinION device. The flowcell was primed with a blank sample created as described above, and left to incubate for 10 min. The priming process was repeated a second time before the sample was loaded. Running MinKNOW version 0.49.2.9 and starting the protocol 'MAP\_48Hr\_Sequencing\_Run.py' initiated the sequencing run. An offline-capable version of MinKNOW, with internet 'ping' disabled and online updates disabled was made available to us by Oxford Nanopore Technologies specifically for the project (available on request from Oxford Nanopore Technologies).

**Data transfer.** With no method of offline analysis available during the outbreak period, there was a dependency on local internet connectivity to facilitate the upload of the raw FAST5 files produced by MinKNOW. A variety of methods were used depending on location and circumstances with the vast majority of the data being uploaded from the European Mobile Laboratories staff accommodation in Coyah, Guinea, via a mobile internet 3G hotspot (TP-LINK M5350 3G hotspot on the MTN mobile network). At times due to unknown factors the upload speed was limited to 2G and took significantly longer. Using Cygwin version 2.0.0 and the Linux tar command a compressed archive containing the first 5,000 to 10,000 .fast5 read files generated by each run was created. This was uploaded to a Google Drive shared directory. Eventually in Coyah we were provided access to a broadband connection (MTN network, 5 Mb s<sup>-1</sup>, established by the World Food Program), which proved to be more reliable than mobile internet.

**Data handling.** Data was downloaded onto a Linux server on the MRC Cloud Infrastructure for Microbial Bioinformatics located in Birmingham, UK. Files were unpacked and basecalled using the Metrichor command-line interface and the workflow 2D Basecalling for MAP-005 (versions 1.14, 1.24 and 1.34). This

software was provided by Oxford Nanopore Technologies (available on request) for the project in order to permit basecalling to be carried out through the Linux command line as part of a pipeline. The MinION generates one direction (1D) and two direction reads (2D). 2D reads are higher quality and were used for analysis. 2D reads that were in the pass filter folder and 2D reads designated as high-quality (due to having more complement events than template events) in the fail folder as determined by poretools were extracted into FASTA (for nanopolish) and FASTQ format (for marginAlign) with poretools version 0.5.1<sup>21</sup>.

**Bioinformatics analysis.** We use a reference mapping approach to detect single nucleotide variants through alignment to a reference strain from early in the outbreak (GenBank accession number EM\_079517)<sup>11</sup>. Due to the nature of the sequencing data, which is dominated by insertion and deletion errors, we do not attempt to call insertion or deletions<sup>14</sup>. Variants were detected using the variants module of the nanopolish software package. Initial nucleotide base alignment was carried out with MarginAlign<sup>12</sup>. Nanopolish then uses the event-level ('squiggle') data generated by the MinION to evaluate candidate variants found in the aligned reads as described in the following section. Variants with a log likelihood ratio of >200 and coverage depth of >50 × (25 × 2D coverage) are accepted and a consensus sequence is generated for each sample. Regions of uncertainty (for example in difficult to sequence homopolymeric regions or primer binding sites), or with low coverage (<50 ×, or 25 × 2D coverage) are masked with an N character. Assuming sufficient genomic coverage is present over a specific amplified variant this approach gives a high true positive variant calling rate. However, failure of individual amplicons to amplify, or unbalanced coverage of regions may reduce this figure. This is assessed, on each individual sample, by artificially mutating the reference genome with 30 randomly chosen mutations. Mutated positions in the references should be detected as variants, using the simplifying assumption that these variants are unlikely to be present in the sample. Any positions not covered by the tiling amplicon scheme (that is, the extreme 5' and 3' ends) are not considered in the true positive rate calculation. Each sample is therefore assigned a quality indicator. Those with a true positive rate (TPR, that is, sensitivity) of ≥75% are included in phylogenetic inferences. Samples with TPR <75% were not used for the phylogenetic analysis presented here.

**Signal-based SNP calling.** SNPs were called using the "variants" module from the nanopolish package (manuscript in preparation, <https://github.com/jts/nanopolish>, branch snp\_calling\_alternative\_models, commit ID 25ea7bac3ab9e1d266079ac105ab2005cfa39a14).

The nanopolish variants program first finds candidate SNPs by finding mismatches between the aligned nanopore reads and the reference genome. These candidate SNPs are clustered into sets of nearby SNPs, an exhaustive set of candidate haplotypes are derived from the possible combinations of SNPs and the haplotype that maximizes the probability of the event-level data called as the sequence for region. We describe each step in detail below.

**Candidate SNP generation.** We iterate over the entire reference genome and examine positions covered by at least 20 nanopore reads. At these well-covered positions we considered any non-reference base that was seen in at least 20% of the nanopore reads to be a candidate SNP. These candidates were passed to the next stage of the pipeline.

**Candidate haplotype generation.** As the MinION sequencer does not measure single bases, but rather current signals dependent on a short sequence of nucleotides that are in the pore, we could not assess each SNP individually. Instead, we partitioned the set of candidate SNPs into groups whose signals may interact and overlap. We determined that SNPs separated by at least 10 bp could be treated independently; therefore we partitioned the candidate SNP set into subsets of SNPs that are within 10 bp of each. For each subset of candidate SNPs we exhaustively generated all possible haplotype sequences by including/excluding the individual SNPs in the subset. As the number of possible combinations of  $n$  SNPs is  $2^n$ , we had to discard subsets that contained more than 10 candidate SNPs or spanned a reference region greater than 100 bp. For each derived haplotype sequence  $S$ , we calculate the likelihood of  $S$  using a modified version of the hidden Markov model (HMM) we previously described<sup>16</sup>.

**Haplotype likelihoods.** The nanopolish HMM calculates the probability of observing a sequence of events emitted by the nanopore, which we denote as  $D$ , given an arbitrary sequence  $S$ . The structure of the HMM is as previously described but now allows events to be "soft-clipped" to better handle uncertainty about where the event-to-sequence alignment starts and ends. In addition, we incorporated a new model from Oxford Nanopore that models the event signals to be dependent on six-base-pair subsequences rather than five-base-pair subsequences. To use this model on SQK-MAP-005 data we calculated a global shift parameter (shift\_offset) that rescales SQK-MAP-005 data to the 6 bp emission functions. We otherwise did not train the emission functions, per-read scaling parameters or transition probabilities of our hidden Markov model.



**Variant calls.** For each subset of candidate SNPs, the haplotype with the largest likelihood is called as the sequence for the region. The SNPs contained on the called haplotype (if any) are output in VCF format. The log likelihood ratio between the called haplotype and the reference haplotype (containing no SNPs) was output as the score for each variant to facilitate downstream filtering. Metadata such as the total depth of the region and the number of reads that support the called haplotype over the reference sequence is also output.

**Validation experiments.** *Dstl amplicons.* Archived Zaire Ebolavirus was amplified using 38 primer pairs, giving approximately 500 base pair amplicons, according to the study protocol. As this work was before in-field sequencing, different versions of the MinKNOW software and Metrichor basecaller were used. Amplicons were sequenced by both MinION and Illumina. An Illumina library was constructed from the same amplicon pool and tagged using the Nextera XT library preparation kit. The library was sequenced on the Illumina MiSeq. Because of the huge excess of coverage generated, this data set was subsampled to 400,000 paired reads before aligning to the EM\_079517 reference sequence using BWA-MEM<sup>22</sup>. After sorting and converting the resulting alignment to BAM using samtools, variants were determined using FreeBayes<sup>23</sup>. A consensus sequence was generated using the vcf2fasta component of vcflib (<https://github.com/ekg/vcflib>). The MinION data was analysed as per the study methods, except for a modification to nanopore to allow it to consider up to 15 variants per segment in order to account for the increased divergence between the genome and the reference. The MinION and Illumina consensus sequences were aligned using the nucmer component of MUMMER and variants determined using the show-snps module<sup>24</sup>. Scripts and documentation for this analysis are in the Github notebook *Dstl validation.ipynb*.

**180 genome analysis.** Six samples of leftover RNA from a previously performed sequencing study<sup>3</sup> were processed at Public Health England Porton Down, as per the methods described in the manuscript. One sample did not yield any sequenceable products, so five genomes (EM\_076534, EM\_076533, EM\_076383, EM\_078416, EM\_076769) were sequenced on MinION at PHE Porton Down. The 11 reaction scheme was used except for sample EM\_076769 when the 19 reaction scheme was used. These sequences were compared with Illumina consensus sequences from the previously published data set in Carroll *et al.*<sup>3</sup>. Variants were identified between the reference genome (EM\_079517) and each of the successfully samples using the show-snps component of MUMMER<sup>24</sup>. Variants detected by our pipeline were compared against expected variants, before and after quality filtering, using custom Python scripts deposited in the Github repository and documented in the IPython Notebook. A phylogeny was inferred using RAxML<sup>25</sup> including the consensus sequences from the validation set along with all of the consensus sequences from Carroll *et al.*<sup>3</sup> MinION sequence accuracy rates for two-direction (2D) reads were determined using A. Quinlan's count-errors.py script (<http://github.com/arq5x/nanopore-scripts>) as described in Quick *et al.*<sup>11</sup>. Scripts and documentation for this analysis are in the Github notebook: *Examine validation runs.ipynb*.

**Analysis of SNP calling sensitivity.** Reads were subsampled at collection time intervals using the poretools times command<sup>21</sup>, simulating the order reads are obtained by real-time sequencing on the nanopore, to demonstrate the effect of coverage on SNP calling sensitivity and log likelihood ratio.

**Analysis of samples from the same patient.** Samples were analysed as part of the real-time surveillance work. The consensus sequences from four pairs of samples each from four individuals were generated. Each pair was compared individually using the show-snps module of MUMMER to investigate differences.

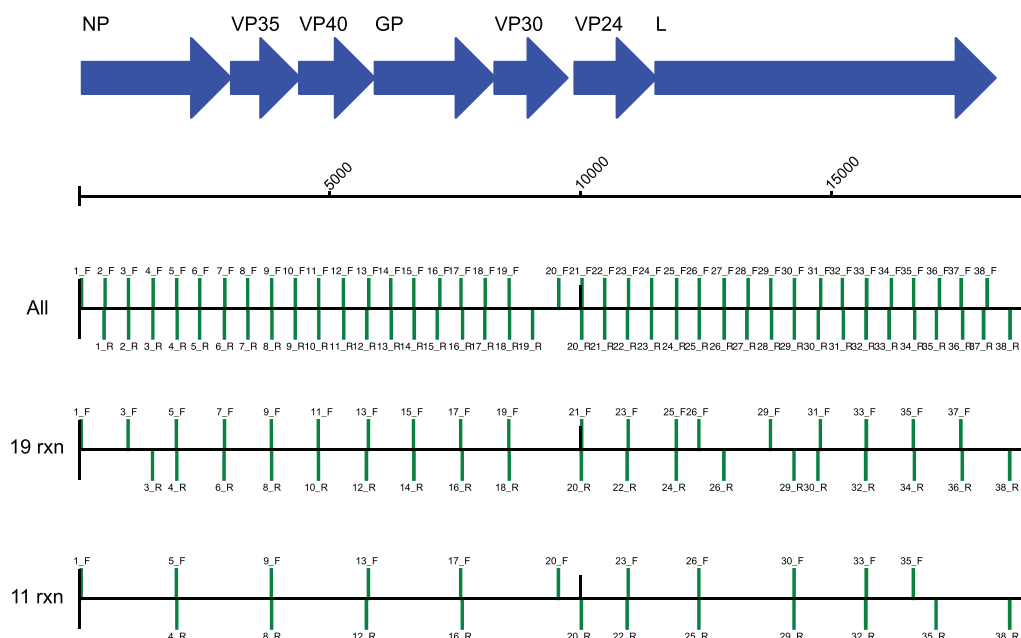
**Detection of putative transmission events from Sierra Leone.** We downloaded the 74 genome sequences made available on <http://virological.org> (<http://virological.org/t/direct-deep-sequencing-in-sierra-leone-yields-73-new-ebov-genomes-from-february-may-2015/134>) and aligned them against sequences from our analysis using MUSCLE<sup>26</sup>. We then generated a phylogenetic tree using FastTree 2 with the GTR model<sup>27</sup>. Any sequences that fell into the GN1 or SL3 lineages were included in future analysis.

**Phylogenetic inferences.** Consensus sequences from real-time sequencing were aligned with previously published genome sequences from Guinea, Sierra Leone and Liberia<sup>7</sup>. To address the over-representation of Sierra Leone sequences in this set we randomly down-sampled available sequences, resulting in a total of 313 sequences from Sierra Leone. Maximum likelihood trees are produced using RAxML 8.2.3 using the GTRGAMMA model and 200 bootstrap replicates<sup>25</sup>. Time-scaled trees were produced with BEAST v1.8.2<sup>28</sup> using a HKY+gamma substitution model<sup>29,30</sup> partitioned by first, second and third codon positions and intergenic regions, a Skygrid tree prior<sup>31</sup> and an uncorrelated lognormal clock<sup>32</sup>, and an uninformative prior on the mean of the molecular clock rate (XML file in the accompanying Github repository). The maximum clade credibility tree was recovered using TreeAnnotator. Phylogenetic trees were annotated using the ete3 Python package.

**Code availability.** Reproducible workflows for the analysis presented here and consensus sequences can be found at <http://github.com/nickloman/ebov> and are freely available under the MIT license. The complete set of bioinformatics scripts are available in a Github repository with associated IPython Notebooks to regenerate the figures and tables presented in this manuscript can be found at <http://github.com/nickloman/ebov>.

20. Smither, S. J. *et al.* Buffer AVL alone does not inactivate Ebola virus in a representative clinical sample type. *J. Clin. Microbiol.* **53**, 3148–3154 (2015).
21. Loman, N. J. & Quinlan, A. R. Poretools: a toolkit for analyzing nanopore sequence data. *Bioinformatics* **30**, 3399–3401 (2014).
22. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at <http://arxiv.org/abs/1303.3997> (2013).
23. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. Preprint at <http://arxiv.org/abs/1207.3907> (2012).
24. Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).
25. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
26. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
27. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).
28. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
29. Hasegawa, M., Kishino, H. & Yano, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160–174 (1985).
30. Yang, Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* **39**, 306–314 (1994).
31. Gill, M. S. *et al.* Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Mol. Biol. Evol.* **30**, 713–724 (2013).
32. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88 (2006).
33. Cock, P. J. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).

A.



B. 19 reactions

Forward	Reverse	Length
1_F	3_R	1426
3_F	4_R	973
5_F	6_R	952
7_F	8_R	941
9_F	10_R	940
11_F	12_R	958
13_F	14_R	906
15_F	16_R	974
17_F	18_R	969
19_F	20_R	1445
21_F	22_R	906
23_F	24_R	958
25_F	26_R	947
26_F	29_R	1898
29_F	30_R	946
31_F	32_R	901
33_F	34_R	963
35_F	36_R	977
37_F	38_R	975

C. 11 reactions

Forward	Reverse	Length
1_F	4_R	1911
5_F	8_R	1901
9_F	12_R	1895
13_F	16_R	1874
17_F	20_R	2406
20_F	22_R	1371
23_F	25_R	1410
26_F	29_R	1898
30_F	32_R	1427
33_F	35_R	1396
35_F	38_R	1921

**Extended Data Figure 1 | Primer schemes employed during the study.** We designed PCR primers to generate amplicons that would span the EBOV genome. **a**, We initially designed 38 primer pairs which were used in the initial validation study and which cover >97% of the EBOV genome. During in-field sequencing we used a 19-reaction scheme or 11-reaction scheme, which generated longer products. The predicted amplicon products are shown with forward primers and reverse primers

indicated by green bars on the forward and reverse strand, respectively, scaled according to the EBOV virus coordinates. **b**, **c**, The amplicon product sizes expected are shown for the 19-reaction scheme (**b**) and the 11-reaction scheme (**c**). No amplicon covers the extreme 3' region of the genome. The last primer pair, 38\_R, ends at position 18578, 381 bases away from the end of the virus genome. The primer diagram was created with Biopython<sup>33</sup>.

## A. Equipment

Item	Number	Model
Thermocycler	1-3	MasterCycler Personal (Eppendorf)
Fluorometer	1	Qubit 3.0 (Life Technologies)
Laptop	2-3	NT310-H (Stone)
MinION	2-3	-
Pipettes	6	P2, 10, 20, 100, 200, 1000 (Gilson)
Microfuge	1-2	
Dry bath	1	Mini Dry Bath Incubator (Starlab)
Magnetic rack	1	MagnaRack (Life Technologies)
Power strip	1	Dependent on country

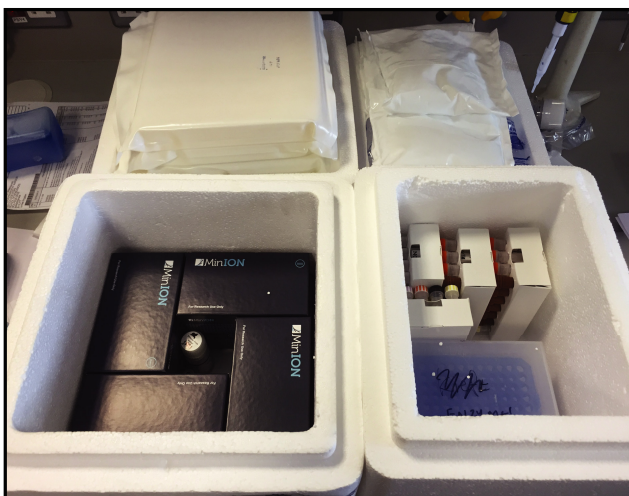


## B. Consumables

Item	Supplier
DNA LoBind Tubes (2 ml)	Eppendorf
Protein LoBind Tubes (2 ml)	Eppendorf
Qubit Assay Tubes	Life Technologies
PCR Tubes with Flat Caps (0.2 ml)	Starlab
Pipette Tips (10 µl, 20 µl, 100 µl, 200 µl, 1000 µl)	Sarstedt
Nitrile Gloves	Kimberly Clark Professional

## C. Reagents

Reagent	Shipping Condition	Supplier
Nuclease-Free Water	Ambient	Qiagen
Ethanol 100%	Ambient	-
HighPrep PCR	Chilled	MAGBIO
Dynabeads His-Tag Isolation and Pulldown	Chilled	Life Technologies
Oligos	Chilled	Sigma
Qubit dsDNA HS Assay Kit	Chilled	Life Technologies
MinION Flowcells	Chilled	Oxford Nanopore Technologies
NEBNext End-Repair Module	Frozen	New England Biolabs
NEBNext dA-Tailing Module	Frozen	New England Biolabs
Blunt/TA Ligase Master Mix	Frozen	New England Biolabs
SuperScript III One-Step RT-PCR System with Platinum Taq DNA Polymerase	Frozen	Life Technologies
SQK-MAP005	Frozen	Oxford Nanopore Technologies



**Extended Data Figure 2 | List of equipment and consumables to establish the genome surveillance system. a–c.** We show the list of equipment (a), disposable consumables (b) and reagents (c) to establish in-field genomic surveillance. Sufficient reagents were shipped for

20 samples. MinION sequencing requires a mix of chilled and frozen reagents. Recommended shipping conditions are specified. The picture underneath depicts MinION flowcells ready for shipping with insulating material (left) and frozen reagents (right).

Base calling	Convert nanopore squiggles to nucleotide sequences	<code>metrichor-cli</code>
Convert to FASTA/FASTQ	Extract basecalled information from nanopore FAST5 files	<code>poretools fasta --type 2D pass/ poretools fasta --type 2D --high- quality fail/</code>
Align to reference	Align sequences to reference	<code>marginAlign --inputModel input.hmm EM_079517.fasta reads.fastq out.sam</code>
Local HMM realignment	Iteratively improve alignment based on nanopore insertion/deletion/substitution rates	
Alignment trimming	Remove alignments outside of primer regions in case of adaptor contamination	<code>align_trim.py</code>
Event alignment	Map individual event k-mers to reference genome guided by base alignment	<code>nanopolish eventalign</code>
Variant calling	Extract candidate mutations from aligned reads, cluster them and evaluate them using a 6-mer HMM	<code>nanopolish variants</code>
Consensus building	Mask positions in the genome with either i) <50x 1-D coverage ii) low-quality variants detected iii) in primer binding site	<code>margin_cons.py</code>

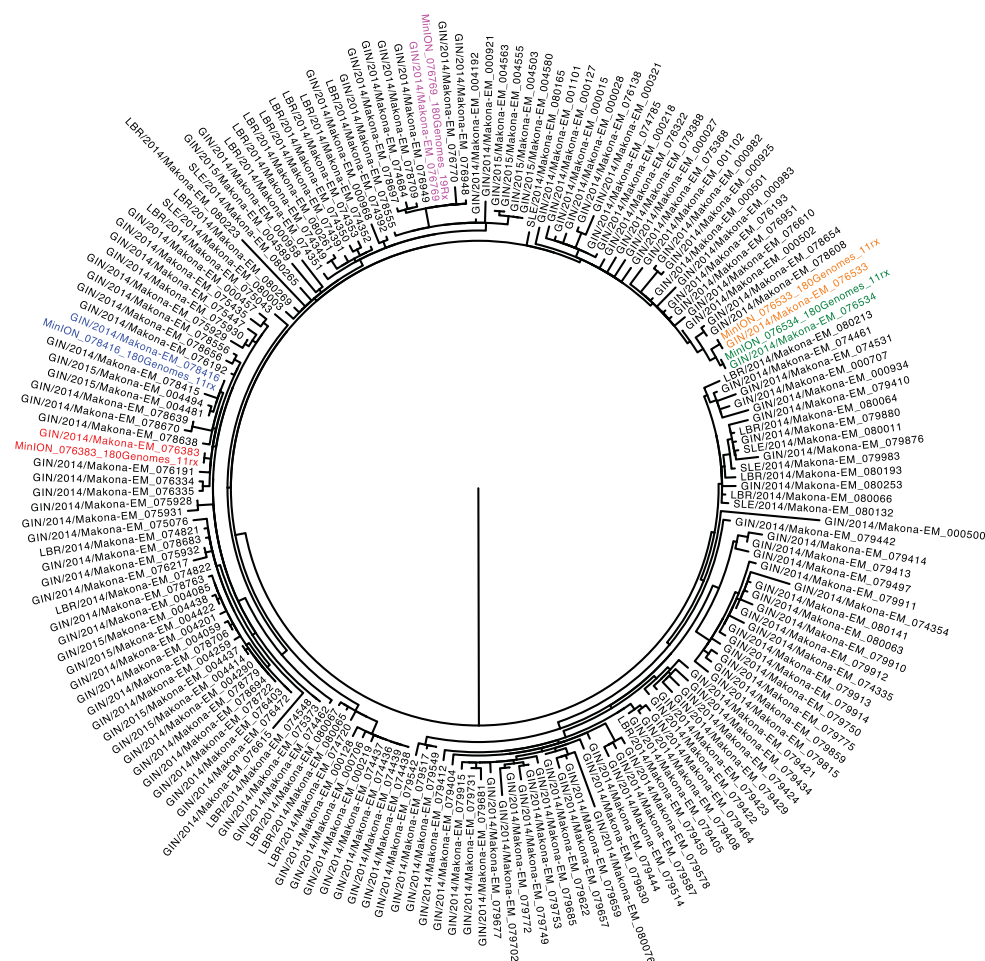
**Extended Data Figure 3 | Bioinformatics workflow.** This figure summarizes the steps performed during bioinformatics analysis (ordered from top to bottom), in order to generate consensus sequences. The right column shows the example software command executed at each step.



A

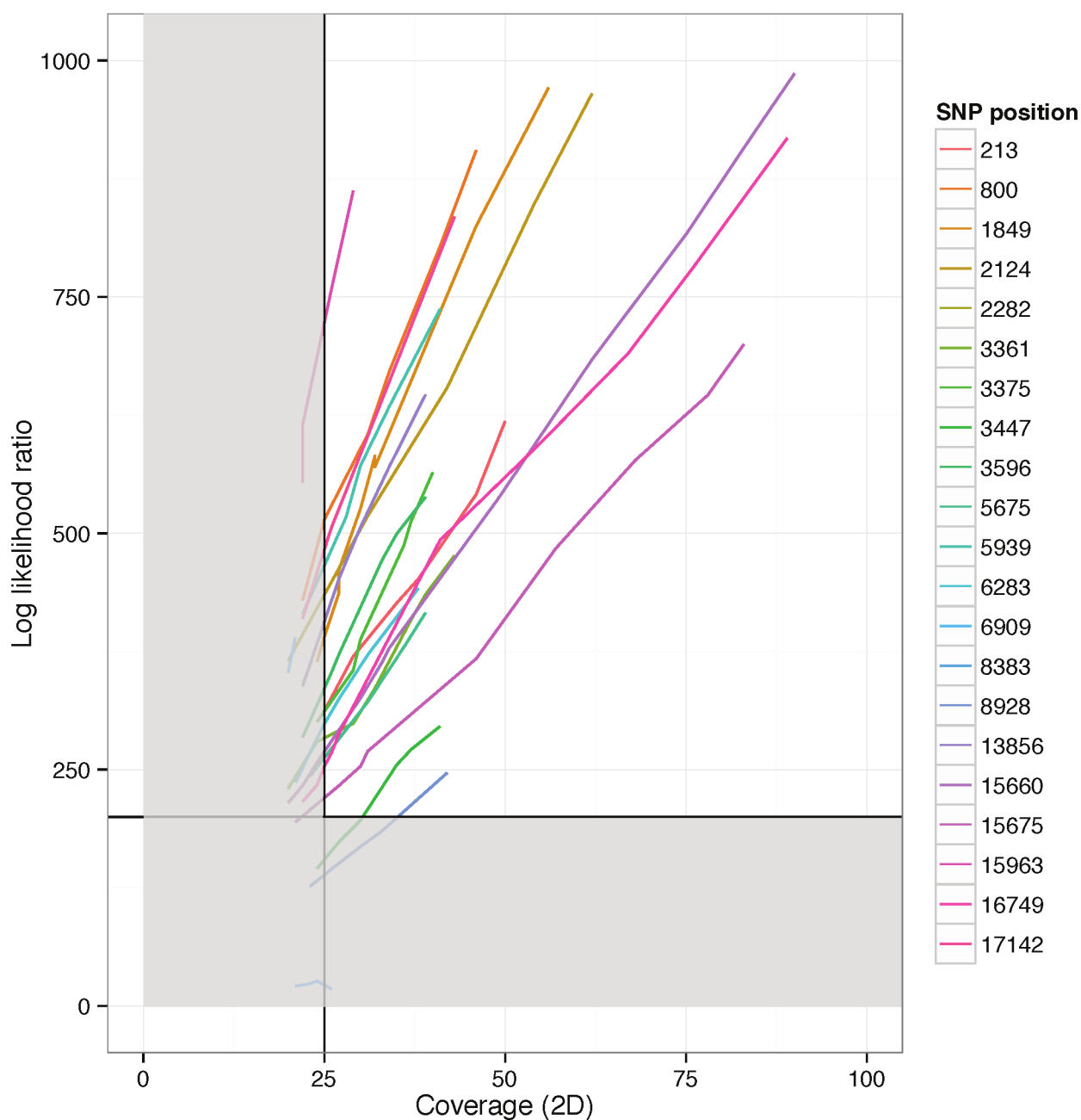
Sample	Reaction scheme	Total mutations	True positives	False positives	False negatives		
					Masked		Wrong
					Outside amplicon scheme	Primer binding site	Erroneous reference call
Pre-quality filtering (all variants)							
076533	11 reactions	20	19	1	1	-	-
076383	11 reactions	18	16	1	1	1	-
078416	11 reactions	18	17	1	-	1	-
076769	19 reactions	19	19	0	-	-	-
Post-quality filtering (high quality variants)							
076533	11 reactions	20	19	0	1	-	-
076383	11 reactions	18	16	0	1	1	-
078416	11 reactions	18	17	0	-	1	-
076769	19 reactions	19	19	0	-	-	-

B



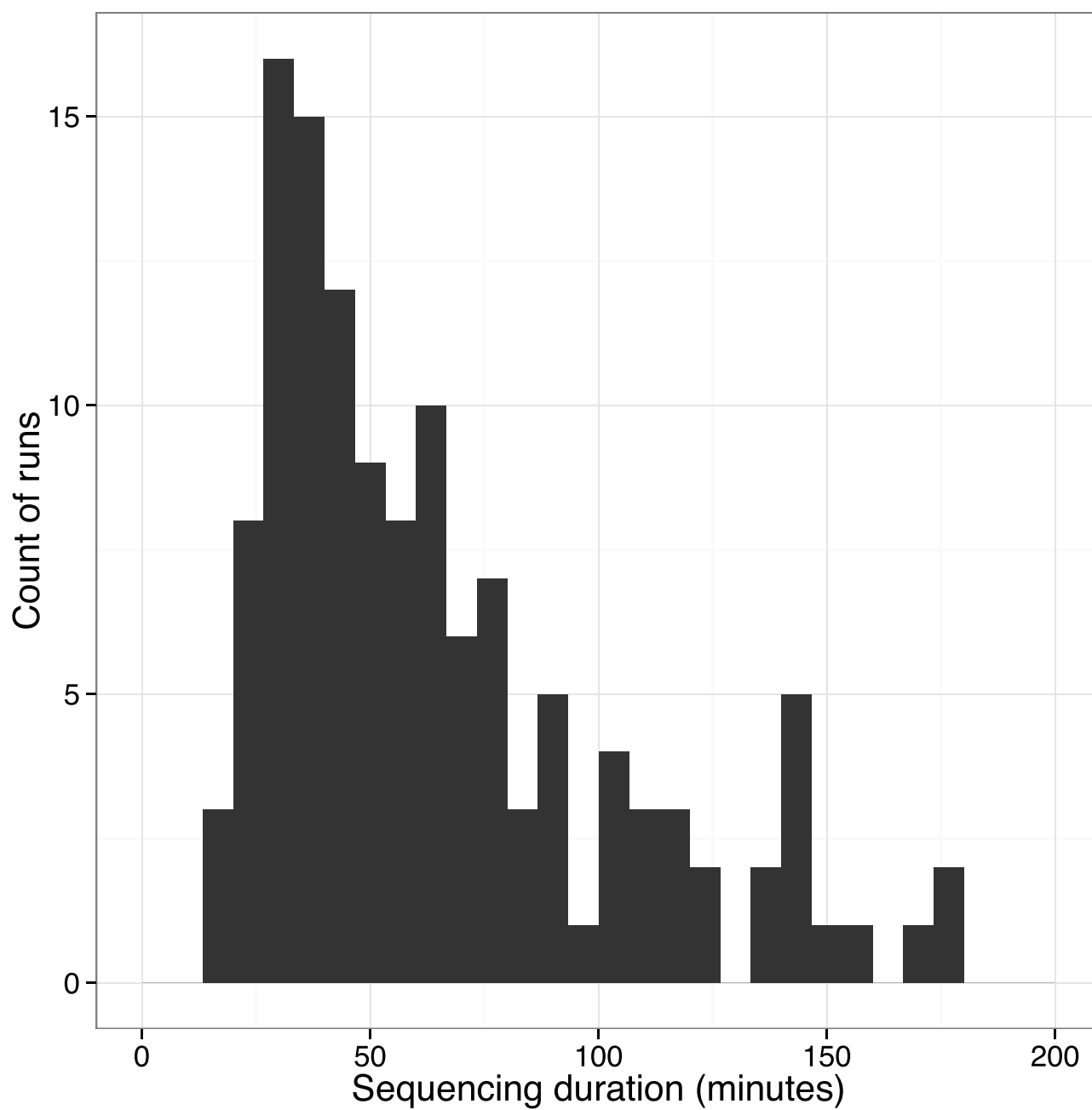
**Extended Data Figure 4 | Results of MinION validation.** **a**, The results of comparing four MinION sequences with Illumina sequences generated as part of a previous study<sup>3</sup> are shown. Each row in the table demonstrates the number of true positives, false positives and false negatives for a sample. False negatives may result in masked sequences, owing to being outside of regions covered by the amplicon scheme, having low coverage or falling within a primer binding site. Results before and after quality filtering (log likelihood ratio of >200) are shown. After quality filtering,

no false positive calls were detected. All detected false negatives were masked with Ns in the final consensus sequence. No positions were called incorrectly. **b**, The four consensus sequences, plus an additional sample that had missing coverage in one amplicon are shown as part of a phylogenetic reconstruction with genomes from Carroll *et al.*<sup>3</sup> Sample labels in red, blue, pink, yellow and blue represent pairs of sequences generated on MinION and Illumina. These fall into identical clusters.

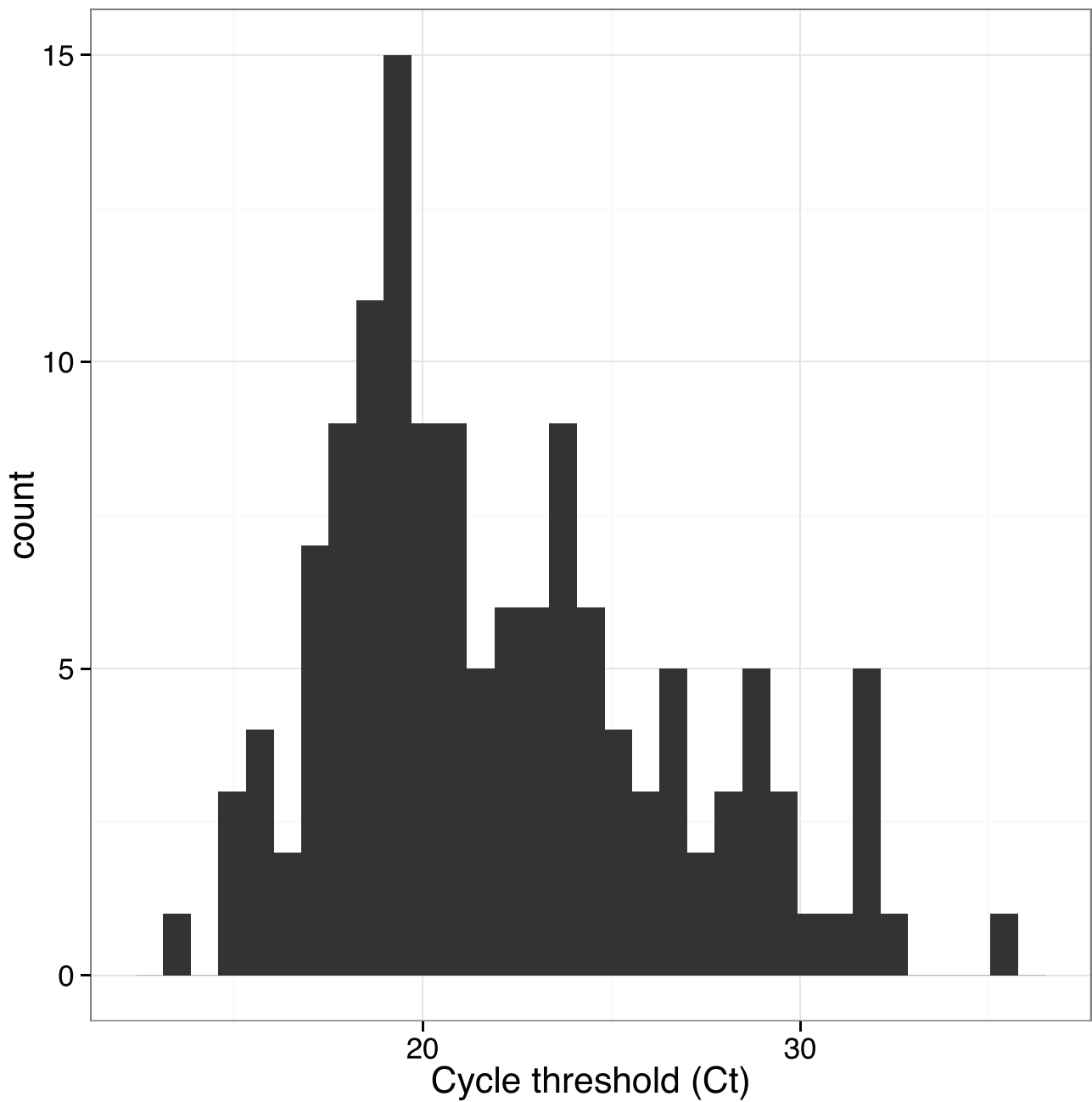


**Extended Data Figure 5 | Relationship between coverage and log-likelihood ratio for sample 076769.** Line-plot showing the relationship between sequence depth of coverage (x axis) and the log likelihood ratio for detected SNPs derived by subsampling reads from a single sequencing run to simulate the effect of low coverage. The horizontal and vertical line indicates the cut-offs (quality and coverage respectively) for consensus

calling. Therefore, all variants are detected below  $25\times$  coverage, and the vast majority meet the threshold quality at  $25\times$  coverage or slightly above. Any combination of log likelihood ratio or coverage that placed variants in the grey box would be represented as a masked position in the final consensus sequence.

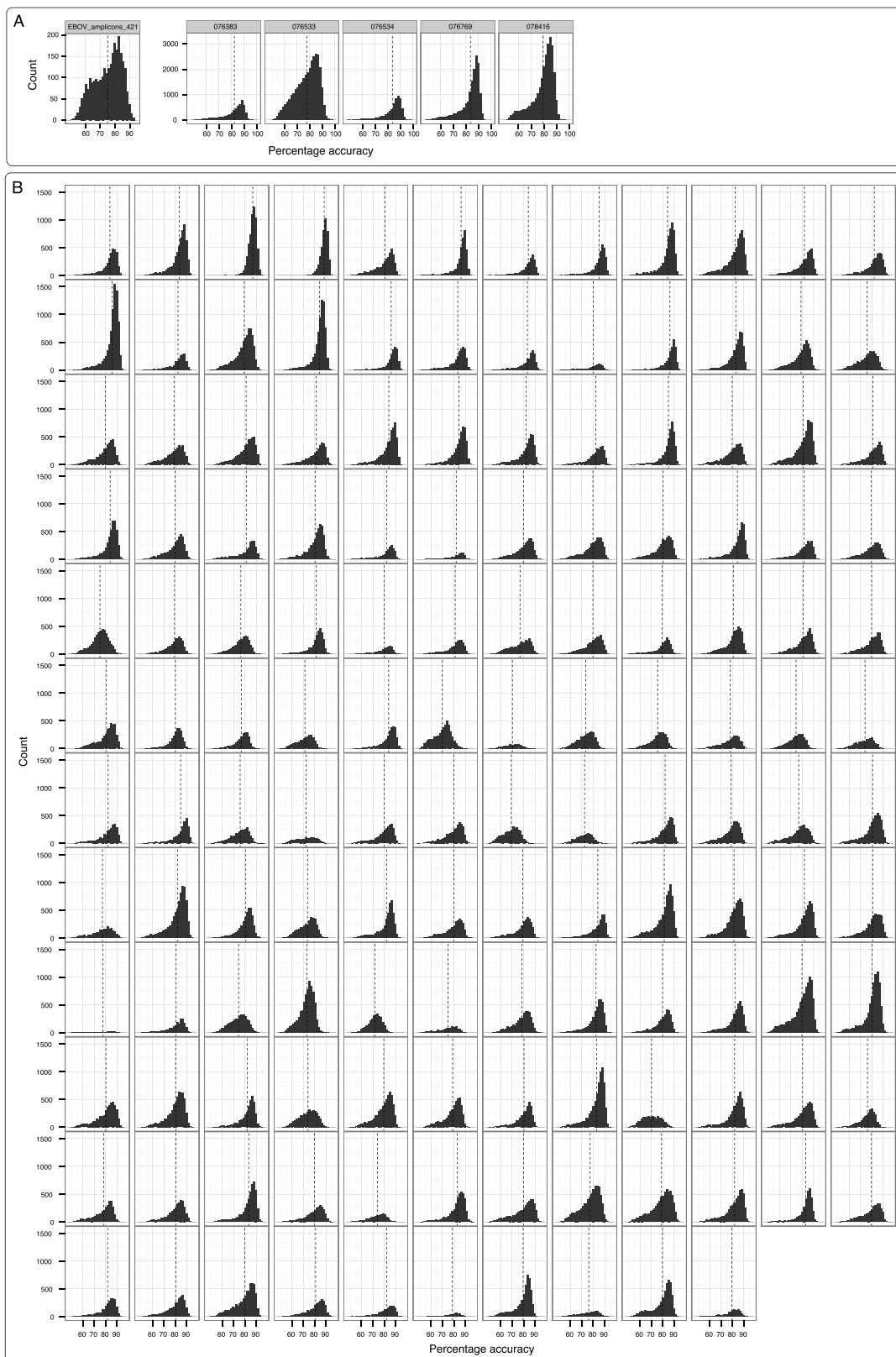


**Extended Data Figure 6 | Duration of MinION sequencing runs.** For each sequence run the sequencing duration, measured as the difference between timestamp of the first read seen and the last read transferred for analysis. 127 runs are shown, with 15 outliers with duration greater than 200 min excluded.



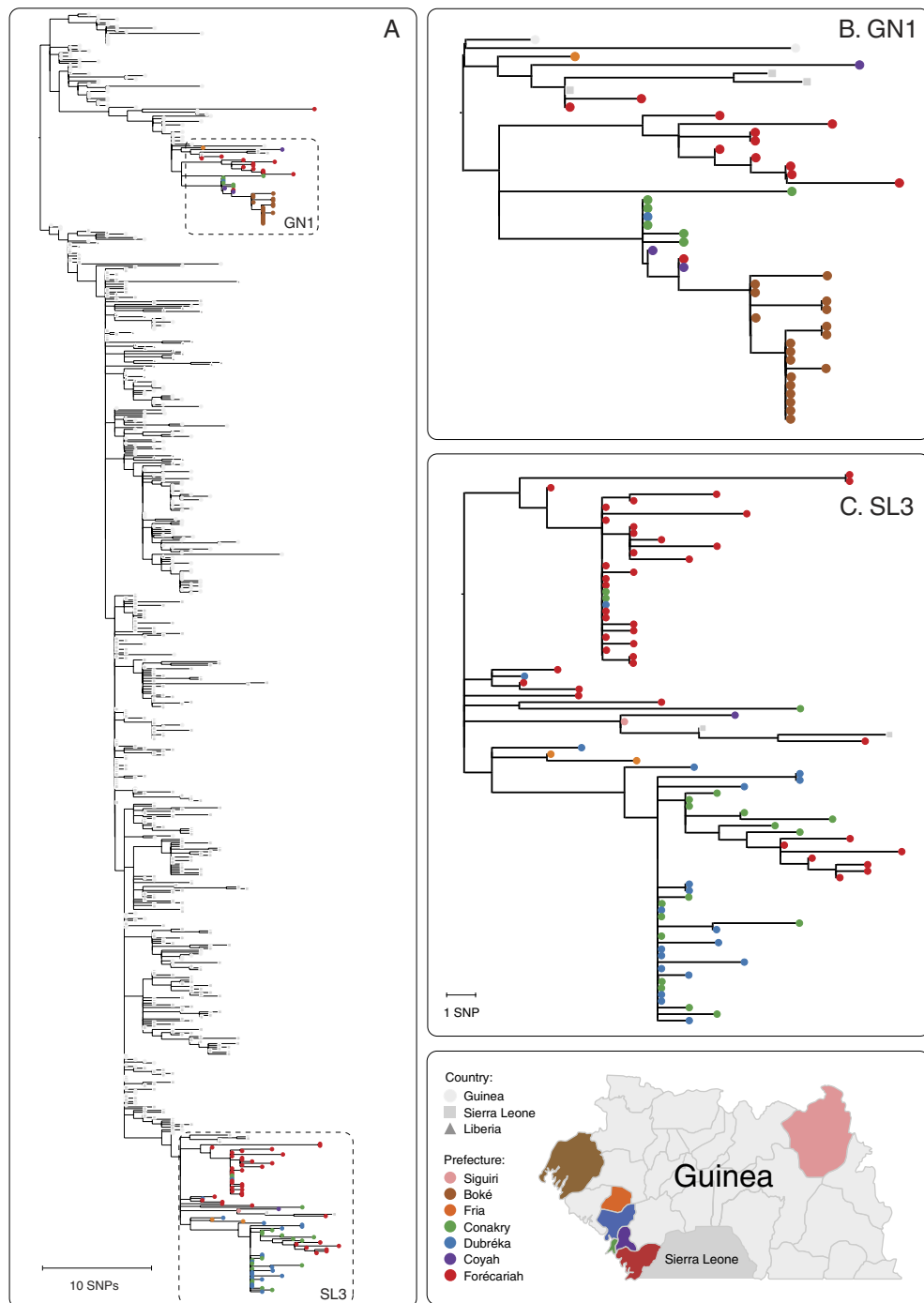
**Extended Data Figure 7 | Histogram of  $C_t$  values for study samples.**  $C_t$  values for samples in the study (where information was available) ranged between 13.8 and 35.7, with a mean of 22.





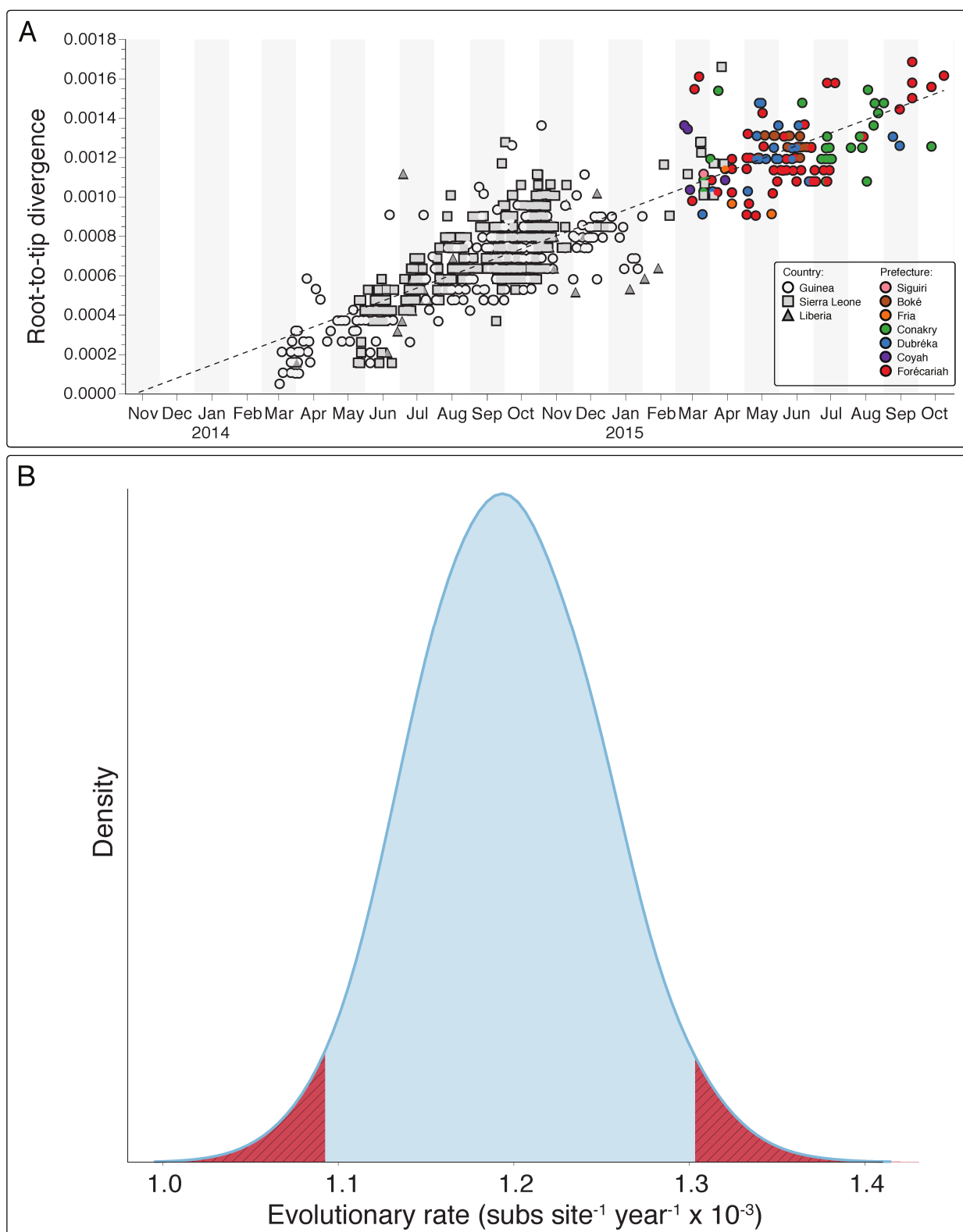
**Extended Data Figure 8 | Sequence accuracy for samples.** a, Accuracy measurements for the entire set of two-direction reads were made for the validation samples, sequenced in the United Kingdom (a) and each of the

142 samples from real-time genomic surveillance (b). Accuracy is defined according to the definition from Quick *et al.*<sup>11</sup>. Vertical dashed lines indicate the mean accuracy for the sample.



**Extended Data Figure 9 | Maximum likelihood phylogenetic inference of 125 Ebola virus samples from this study with 603 previously published sequences.** Coloured nodes are from this study. Node shape reflects country of origin. **a–c**, the entire data set is shown (**a**), with

zoomed regions focusing on lineages GN1 (**b**) and SL3 (**c**) identified during real-time sequencing. Map figure adapted from SimpleMaps website (<http://simplemaps.com/resources/svg-gn>).



**Extended Data Figure 10 | Root-to-tip divergence plot and mean evolutionary rate estimate.** **a**, Root-to-tip divergence plot for the 728 Ebola samples generated through maximum likelihood analysis. Samples from real-time genomic surveillance are coloured as per Fig. 3 and Extended Data Fig. 9. **b**, Mean evolutionary rate estimate (in substitutions

per site per year) across the EBOV phylogeny recovered using BEAST under a relaxed lognormal molecular clock. Blue area corresponds to the 95% highest posterior density (HPD) (mean of the distribution is  $1.19 \times 10^{-3}$ , 95% HPDs:  $1.09$ – $1.29 \times 10^{-3}$  substitutions per site per year). Hatched regions in red are outside the 95% HPD intervals.

# Structure- and function-based design of *Plasmodium*-selective proteasome inhibitors

Hao Li<sup>1,2†</sup>, Anthony J. O'Donoghue<sup>3†</sup>, Wouter A. van der Linden<sup>1</sup>, Stanley C. Xie<sup>4</sup>, Euna Yoo<sup>2</sup>, Ian T. Foe<sup>2</sup>, Leann Tilley<sup>4</sup>, Charles S. Craik<sup>3</sup>, Paula C. A. da Fonseca<sup>5</sup> & Matthew Bogoy<sup>2</sup>

The proteasome is a multi-component protease complex responsible for regulating key processes such as the cell cycle and antigen presentation<sup>1</sup>. Compounds that target the proteasome are potentially valuable tools for the treatment of pathogens that depend on proteasome function for survival and replication. In particular, proteasome inhibitors have been shown to be toxic for the malaria parasite *Plasmodium falciparum* at all stages of its life cycle<sup>2–5</sup>. Most compounds that have been tested against the parasite also inhibit the mammalian proteasome, resulting in toxicity that precludes their use as therapeutic agents<sup>2,6</sup>. Therefore, better definition of the substrate specificity and structural properties of the *Plasmodium* proteasome could enable the development of compounds with sufficient selectivity to allow their use as anti-malarial agents. To accomplish this goal, here we use a substrate profiling method to uncover differences in the specificities of the human and *P. falciparum* proteasome. We design inhibitors based on amino-acid preferences specific to the parasite proteasome, and find that they preferentially inhibit the  $\beta 2$ -subunit. We determine the structure of the *P. falciparum* 20S proteasome bound to the inhibitor using cryo-electron microscopy and single-particle analysis, to a resolution of 3.6 Å. These data reveal the unusually open *P. falciparum*  $\beta 2$  active site and provide valuable information about active-site architecture that can be used to further refine inhibitor design. Furthermore, consistent with the recent finding that the proteasome is important for stress pathways associated with resistance of artemisinin family anti-malarials<sup>7,8</sup>, we observe growth inhibition synergism with low doses of this  $\beta 2$ -selective inhibitor in artemisinin-sensitive and -resistant parasites. Finally, we demonstrate that a parasite-selective inhibitor could be used to attenuate parasite growth *in vivo* without appreciable toxicity to the host. Thus, the *Plasmodium* proteasome is a chemically tractable target that could be exploited by next-generation anti-malarial agents.

The *Plasmodium* proteasome represents a viable target for anti-malarial drugs owing to its essential nature in protein turnover and the parasite's need to rapidly divide inside host cells<sup>9</sup>. We have previously shown that differences exist in the reliance of human cells and *P. falciparum* on specific proteasome activities for survival<sup>10</sup>. Therefore, we reasoned that a structural and biochemical understanding of this enzyme complex could yield compounds with desirable safety profiles due to their selectivity for the parasite proteasome. We purified the *P. falciparum* 20S proteasome<sup>11</sup>, activated it with human PA28 $\alpha$ <sup>12,13</sup> (Extended Data Fig. 1a), and determined the substrate preferences of this activated proteasome by monitoring the degradation pattern of 228 diverse synthetic tetradecapeptides using liquid chromatography–tandem mass spectrometry<sup>14,15</sup>. A frequency profile was generated using iceLogo software<sup>16</sup>, indicating which amino acids were most and

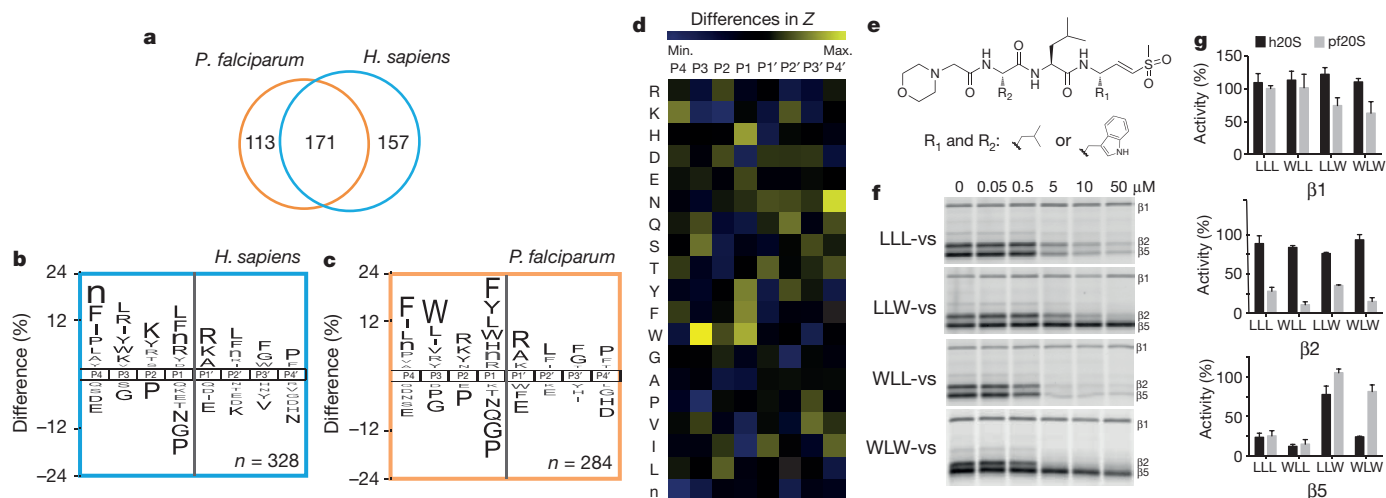
least favoured in the subsites surrounding each cleaved bond. In parallel, we generated a substrate specificity profile for the human 20S/PA28 $\alpha$  proteasome under identical conditions. We detected 284 and 328 cleavage sites for the activated *P. falciparum* and human 20S proteasomes, respectively, with 171 sites being cleaved by both (Fig. 1a). However, 113 and 157 sites were uniquely cleaved by *P. falciparum* and human proteasomes, respectively (Fig. 1a–c and Extended Data Fig. 1b, c), with the major differences occurring on the amino (N)-terminal side of the cleavage site, namely at the P1 and P3 sites. For the P1 position which is directly adjacent to the cleavage site, the human proteasome showed chymotrypsin-like (Leu/Phe/Tyr), trypsin-like (Arg), and caspase-like (Asp) activity (Fig. 1b), consistent with previous studies<sup>17</sup>, while the *P. falciparum* proteasome showed strong preference for aromatic residues (Fig. 1c). Direct comparison of the specificity profiles of both proteasomes revealed a clear preference for Trp at P1 and P3 in the parasite proteasome (Fig. 1d and Extended Data Fig. 1b, c).

Using the canonical tri-leucine scaffold found in common proteasome inhibitors such as MG132 and Z-L3-VS<sup>18,19</sup>, we systematically replaced the Leu residues at the P1 and P3 positions with Trp, resulting in the compounds LLW-vs, WLL-vs and WLW-vs (Fig. 1e and Extended Data Fig. 2a). We treated purified *P. falciparum* 20S proteasomes with each compound and measured residual activity of the catalytic sites using an activity-based probe<sup>10</sup> (Fig. 1f). Changing the P1 position to Trp (LLW-vs) markedly reduced inhibition of the  $\beta 5$  active site, without altering the activity towards the  $\beta 2$ -subunit, while Trp in the P3 position (WLL-vs) resulted in potent inhibition of both  $\beta 2$  and  $\beta 5$  activities. Finally, Trp substitution in both P1 and P3 positions (WLW-vs) resulted in a potent inhibitor that was highly selective for the parasite proteasome  $\beta 2$ -subunit. With the exception of LLW-vs, these compounds inhibited the human  $\beta 5$ -subunit while all compounds only weakly inhibited the human  $\beta 1$  and  $\beta 2$  activities (Fig. 1g and Extended Data Figs 1d and 2b). Therefore, these inhibitors primarily differ in their selectivity towards the  $\beta 2$ -subunits and this selectivity was retained in human and *P. falciparum* inhibitor treatments in live culture (Extended Data Fig. 2c).

We investigated the structural basis for the  $\beta 2$  selectivity of the *P. falciparum* 20S proteasome towards WLW-vs by high-resolution cryo-electron microscopy (cryo-EM) and single-particle analysis, using a strategy similar to that which we previously used to determine the structure of an inhibitor-bound human 20S proteasome<sup>20</sup>. The resolution of the final map was estimated at 3.6 Å (Fig. 2 and Extended Data Fig. 3a–e), with local resolution decreasing for solvent exposed surfaces (Fig. 2a, b). In our map, sheet-forming  $\beta$ -strands are well resolved (Fig. 2d) and side chains are well defined for both  $\beta$ -strands and  $\alpha$ -helices (Fig. 2e, f), particularly when stabilized by intra-protein contacts or steric restraints. We were able to unambiguously assign a

<sup>1</sup>Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, California 94305, USA. <sup>2</sup>Department of Pathology, Stanford University School of Medicine, Stanford, California 94305, USA. <sup>3</sup>Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, California 94158, USA. <sup>4</sup>Department of Biochemistry and Molecular Biology, Bio21 Institute, University of Melbourne, Melbourne 3010, Victoria, Australia. <sup>5</sup>MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH, UK. <sup>†</sup>Present addresses: Molecular Engineering Laboratory, Biomedical Sciences Institute, Agency for Science, Technology and Research, Singapore 138673 (H.L.); Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, California 92093, USA (A.J.O.).





**Figure 1 | Substrate profile of the activated human and *P. falciparum* 20S proteasome guides inhibitor design.** **a**, Total number of cleavage sites detected after incubation for 4 h of the activated human and *P. falciparum* proteasome with the peptide library. **b**, **c**, The iceLogos generated from the cleavages are shown for human (**b**) and *P. falciparum* (**c**) proteasome. Amino acids that are most and least favoured at each position are shown above and below the axis, respectively. Lower-case 'n', norleucine; amino acids in black text are statistically significant ( $P < 0.05$ , unpaired two-tailed Student's *t*-test). **d**, The Z-score for amino acids at each position (P4–P4') was calculated for both the human and the parasite proteasome on the basis of the cleavages in **a**, and the difference between the Z-scores

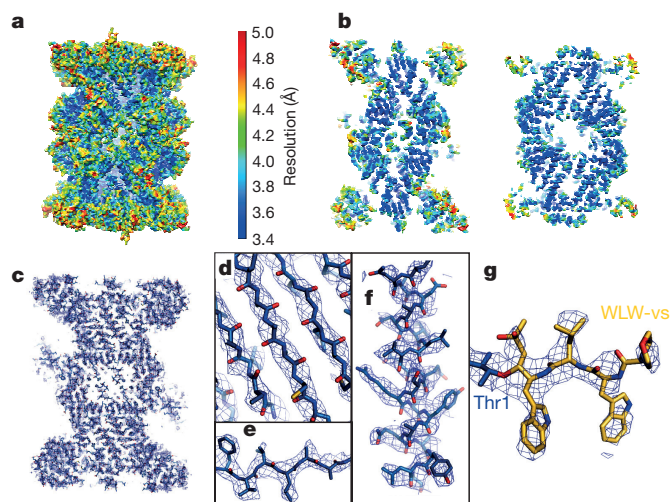
is shown as a heatmap. **e**, Inhibitors are designed by substituting Trp at either the P1 and/or P3 positions in the morpholino-capped tri-leucine vinyl sulfone. Residue 1 ( $R_1$ ) and 2 ( $R_2$ ) refer to the amino acids Leu or Trp. **f**, Inhibition of purified *P. falciparum* 20S as assessed by activity-based probe labelling. The same experiment was repeated for the human 20S proteasome (Extended Data Fig. 2b). **g**, Activity of each subunit in human or *P. falciparum* proteasome after 10  $\mu$ M inhibitor treatment was determined by image quantification of the intensity of probe labelling and normalized to mock-treated control. Error bars, s.d.  $n = 3$  purified proteasome from three independent experiments (for gel source data, see Supplementary Fig. 1a, b).

clear density extending from the  $\beta 2$ -subunit N-terminal threonine to the WLW-vs inhibitor, with the vinyl sulfone group and the WLW side chains clearly resolved in a  $\beta$ -strand secondary structural conformation (Fig. 2g).

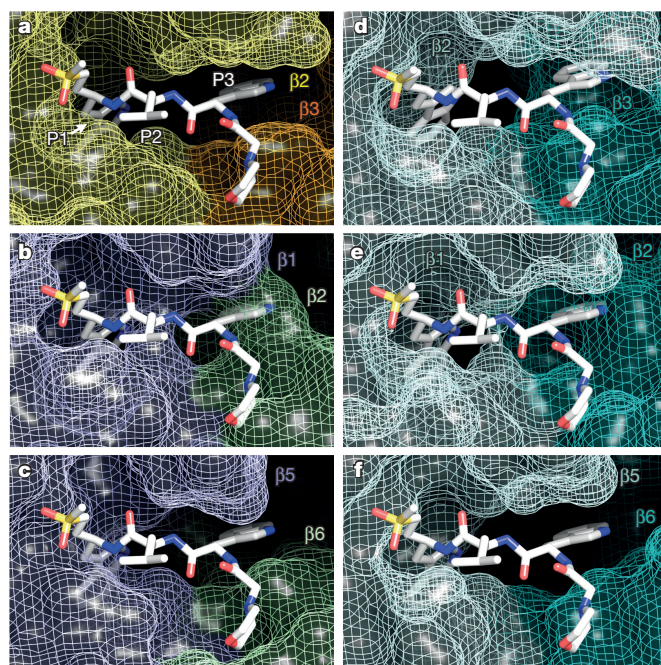
When the protein coordinates for the *P. falciparum*  $\beta 1$ - and  $\beta 5$ -subunits are aligned to those of the ligand-bound  $\beta 2$ -subunit, the resulting superimposition of the WLW-vs ligand into the  $\beta 1$  and  $\beta 5$  active sites clearly reveals space constraints for its binding. While the more open conformation of the  $\beta 2$  binding pocket easily accommodates the WLW-vs side chains (Fig. 3a), the ligand overlay into the  $\beta 1$  and  $\beta 5$  active sites results in the burying of its Trp side chains into the protein van der Waals surface (Fig. 3b, c), with the exception of the P3 position in the  $\beta 5$  active site (Fig. 3c). We extended this comparison to the protein coordinates of the human 20S proteasome active sites, using the coordinates of the human apo complex<sup>21</sup> (Protein Data Bank (PDB) accession 4R3O). Superimposing the WLW-vs inhibitor onto the human apo proteasome core (Fig. 3d–f) shows its active sites have no accessibility for the P1 and P3 Trp side chains of the inhibitor, with the exception of the P3 position at the  $\beta 5$  active site (Fig. 3f). Similar observations can be made when using the protein model of the human proteasome core bound to the inhibitor AdaAhx<sub>3</sub>-LLL-vs (PDB 5A0Q) (Extended Data Fig. 3f–h). This structural comparison of the accessibility of *P. falciparum* and human proteasome active sites is consistent and explains the observed specificity of the WLW-vs inhibitor towards the *P. falciparum*  $\beta 2$ -subunit, which would be difficult to predict without direct high-resolution structural information. The results presented here show an important and novel application of high-resolution cryo-EM to interpret the basis for the specific binding of a ligand to an active site, where the information obtained can be harnessed to guide further ligand development for therapeutic usage.

*P. falciparum* parasites can become resistant to killing by the front-line artemisinin (ART) drug family<sup>22–24</sup>; however, these ART-treated parasites are particularly sensitive to disruption of normal proteasome function<sup>25</sup>. We previously showed that inhibition of the proteasome  $\beta 5$ -subunit is required for attenuation of *P. falciparum* growth<sup>10</sup>. However, we reasoned that a sublethal dose of the *Plasmodium*

$\beta 2$ -selective inhibitor WLW-vs may still be sufficient to enhance killing by ART in both sensitive and resistant parasites. When laboratory-adapted clinical isolates were treated at the early ring stage of the intraerythrocytic cycle, ART-resistant parasites (PL7) showed close to twofold higher sensitivity to WLW-vs than ART-sensitive parasites (PL2) (Fig. 4a). In addition, we observed synergism between dihydroartemisinin (DHA) and WLW-vs at concentrations where WLW-vs selectively inhibits the  $\beta 2$ -subunit of the *Plasmodium* proteasome (Fig. 4b and



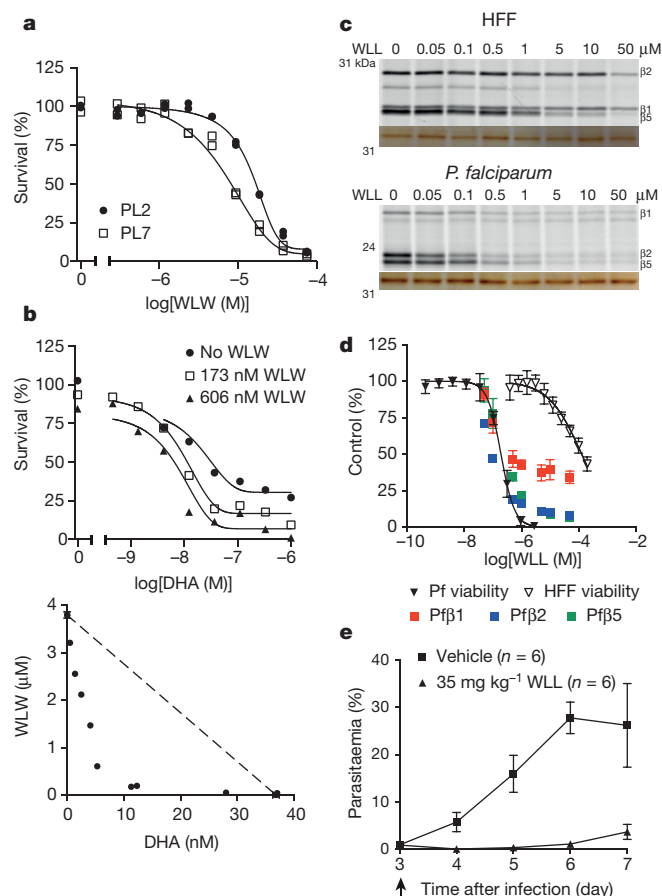
**Figure 2 | Structure of the *P. falciparum* 20S proteasome core bound to the inhibitor WLW-vs, determined by cryo-EM and single-particle analysis.** **a**, The cryo-EM map is surface rendered and colour coded according to local resolution as determined by ResMap. **b**, Sections of the map as shown in **a**, revealing the higher resolution at the internal regions of the complex, compared with solvent exposed surfaces. **c**, Section of the cryo-EM map, as mesh representation, showing the overall fitting of the protein model. **d**–**g**, Detailed views of the map showing clear separation of sheet-forming  $\beta$ -strands (**d**), good recovery of side chains, both in  $\beta$ -strands (**e**) and  $\alpha$ -helices (**f**), and the densities for the WLW-vs inhibitor at the proteasome  $\beta 2$  active site (**g**).



**Figure 3 | Structural comparison of the *P. falciparum* and human proteasome 20S core active sites.** **a**, Coordinates of the *P. falciparum* proteasome  $\beta 2$  active site bound to WLW-vs. The inhibitor P1, P2 and P3 positions are indicated. **b**, **c**, Coordinates of the *P. falciparum* proteasome  $\beta 1$  (**b**) and  $\beta 5$  (**c**) active sites with superimposed coordinates of the WLW-vs inhibitor, as shown in **a**. **d–f**, Coordinates of the apo human proteasome 20S core (PDB 4R3O)  $\beta 2$  (**d**),  $\beta 1$  (**e**), and  $\beta 5$  (**f**) active sites, with superimposed coordinates of the WLW-vs inhibitor, as shown in **a**. In all parts the protein is represented as van der Waals surfaces and the inhibitor as sticks.

Extended Data Fig. 4a). As parasites are normally insensitive to a pulse inhibition of the proteasome  $\beta 2$  active site alone<sup>10</sup>, this emphasizes the heightened sensitivity of *P. falciparum* to proteasome inhibition in the presence of DHA.

We have previously shown that co-inhibition of the *Plasmodium* proteasome  $\beta 2$ - and  $\beta 5$ -subunits results in complete parasite growth attenuation at all stages of its blood cycle, even after short inhibitor treatment times<sup>10</sup>. WLL-vs is a potent inhibitor of the parasite  $\beta 2$  and  $\beta 5$  active sites, while only effective against the human  $\beta 5$  activity (Figs 1g, 3 and Extended Data Fig. 1d, e). WLL-vs is therefore likely to have potent anti-malarial activity with minimal toxicity to the host, since a short inhibition of host  $\beta 5$  activity is well tolerated in non-transformed mammalian cells<sup>26</sup>. We treated live *P. falciparum* cultures and primary human fibroblasts for 1 h with WLL-vs, and were surprised to find that WLL-vs inhibited all three proteasome catalytic sites in *P. falciparum*, while having little effect on any of the human subunits until high micromolar concentrations (Fig. 4c). This increased potency is probably due to inhibitor accumulation in the intact parasites and enhanced inhibition of the parasite 26S proteasome (20S with regulatory subunit 19S proteasome complex)<sup>27</sup>. This translated into a large therapeutic window for the WLL-vs inhibitor in 1 h and 72 h viability assays (Fig. 4d and Extended Data Fig. 1e). Although the *P. falciparum* proteasome inhibition corresponds well to the decrease in parasite survival (Fig. 4d), we sought to determine whether off-target toxicity was occurring. First, a diastereomer of WLL-vs containing a P1 D-Leu (WL(d)L-vs) showed a dramatic reduction in compound potency (Extended Data Fig. 5a–c), indicating that these compounds were not indiscriminately cytotoxic. Second, vinyl-sulfones can react with cysteine proteases<sup>28</sup>; therefore, we assessed reactivity of WLL-vs towards *P. falciparum* cysteine proteases using the activity-based probe BODIPY-TMR-DCG04 (ref. 29). We observed inhibition of cysteine proteases with WLL-vs only at concentrations that were



**Figure 4 | Exploiting differences in the  $\beta 2$ -subunits of the two proteasome species for selective parasite killing.** **a**, **b**,  $\beta 2$ -subunit-selective inhibitor WLW-vs can synergize with DHA treatment in ART-resistant *P. falciparum*. **a**, Viability 72 h after treatment for 1 h of ART-sensitive (PL2) and -resistant (PL7) parasites at early ring stage (2 h after invasion) with WLW-vs (abbreviated as WLW). **b**, Top, dose-response to DHA in the absence and presence of sublethal doses of WLW-vs. Bottom, isobologram of the two inhibitors at the 50% lethal dose. PL7 parasites were treated for 3 h at early ring stage (4 h after invasion). **a**, **b**, Data shown are individual data points (two technical replicates for **a**), representative of three independent experiments. **c–e**, WLL-vs co-inhibits  $\beta 2$  along with the other two catalytic subunits, resulting in potent parasite killing. **c**, Human foreskin fibroblasts (HFFs) or *P. falciparum* schizonts were treated with WLL-vs for 1 h, and activity-based probe labelling was performed post-lysis (for gel source data, see Supplementary Fig. 1c, d). Fluorescent gel scan (top) and silver stain (bottom) are shown. **d**, Non-confluent, replicating HFFs ( $n = 9$  cell cultures, from three independent experiments of triplicates) or *P. falciparum* at ring stage ( $n = 6$  parasite cultures, from two independent experiments of triplicates) were pulse-treated for 1 h with WLL-vs and viability was determined after 71 h. Activities of the *P. falciparum* proteasome subunits after inhibitor treatment were determined as described in Methods ( $n = 3$  parasite lysates from three independent experiments). Error bars, s.d. **e**, Balb/c female mice were infected with *P. chabaudi* and treated with a single intravenous dose of vehicle ( $n = 6$  mice) or 35 mg kg<sup>-1</sup> WLL-vs ( $n = 6$  mice) on day 3 after infection (arrow). Parasitaemia and weight of the mice were monitored daily until the infection was resolved naturally by the host (day 7 onwards). Error bars, s.d.

>tenfold over the 1 h treated half-maximum effective concentration (EC<sub>50</sub>) (Extended Data Fig. 5d). Third, we assessed parasite morphology 24 h after a 1 h WLL-vs pulse treatment and observed arrested parasite development and eventual parasite death at high inhibitor concentrations (Extended Data Fig. 5e). The absence of a food vacuole defect indicated that the falcipains in this digestive organelle were not inhibited. Taken together, these data confirm that WLL-vs exerts its anti-malarial effects through direct proteasome inhibition.



Given the large therapeutic window of WLL-vs, we assessed its efficacy in a rodent *Plasmodium chabaudi* infection model. We infected Balb/c female mice with  $1 \times 10^6$  parasites and treated them 3 days later using a single bolus dose of 35 mg kg<sup>-1</sup> WLL-vs or vehicle, by tail vein injection. We observed almost complete reduction of parasite burden after WLL-vs treatment (Fig. 4e) and good tolerance to the drug as assessed by appearance, activity, and weight retention (Extended Data Fig. 4b). The efficacy of this treatment was further confirmed in an intraperitoneal single bolus dosing study of mice infected with *P. chabaudi* (Extended Data Fig. 4c).

In summary, we have shown through biochemical and structural studies that the *Plasmodium* proteasome is sufficiently unique from the human proteasome for selective targeting. Furthermore we show that even sublethal doses of proteasome inhibitor can lead to parasite sensitization to ART. Our lead molecule, WLL-vs, is a mechanism-based peptidic inhibitor that is chemically similar to the US Food and Drug Administration-approved drug Kyprolis. We recognize that the unique disease epidemiology of malaria calls for stringent criteria in drug safety, stability, and delivery, and we are currently optimizing WLL-vs to generate a class of compounds with improved potency and oral bioavailability. Previous success in optimizing peptide-based agents<sup>30</sup> for efficient oral delivery offers promise in this class of inhibitors as next-generation anti-malarials.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 10 July; accepted 18 December 2015.**

- Voges, D., Zwickl, P. & Baumeister, W. The 26S proteasome: a molecular machine designed for controlled proteolysis. *Annu. Rev. Biochem.* **68**, 1015–1068 (1999).
- Gantt, S. M. *et al.* Proteasome inhibitors block development of *Plasmodium* spp. *Antimicrob. Agents Chemother.* **42**, 2731–2738 (1998).
- Czesny, B., Goshu, S., Cook, J. L. & Williamson, K. C. The proteasome inhibitor epoxomicin has potent *Plasmodium falciparum* gametocytocidal activity. *Antimicrob. Agents Chemother.* **53**, 4080–4085 (2009).
- Kreidenweiss, A., Krenmsner, P. G. & Mordmüller, B. Comprehensive study of proteasome inhibitors against *Plasmodium falciparum* laboratory strains and field isolates from Gabon. *Malar. J.* **7**, 187 (2008).
- Prudhomme, J. *et al.* Marine actinomycetes: a new source of compounds against the human malaria parasite. *PLoS ONE* **3**, e2335 (2008).
- Prasad, R. *et al.* Blocking *Plasmodium falciparum* development via dual inhibition of hemoglobin degradation and the ubiquitin proteasome system by MG132. *PLoS ONE* **8**, e73530 (2013).
- Mok, S. *et al.* Population transcriptomics of human malaria parasites reveals the mechanism of artemisinin resistance. *Science* **347**, 431–435 (2015).
- Mbengue, A. *et al.* A molecular mechanism of artemisinin resistance in *Plasmodium falciparum* malaria. *Nature* **520**, 683–687 (2015).
- Aminake, M. N., Arndt, H. D. & Pradel, G. The proteasome of malaria parasites: a multi-stage drug target for chemotherapeutic intervention? *Int. J. Parasitol.* **42**, 1–10 (2012).
- Li, H. *et al.* Assessing subunit dependency of the *Plasmodium* proteasome using small molecule inhibitors and active site probes. *ACS Chem. Biol.* **9**, 1869–1876 (2014).
- Li, H. *et al.* Validation of the proteasome as a therapeutic target in *Plasmodium* using an epoxyketone inhibitor with parasite-specific toxicity. *Chem. Biol.* **19**, 1535–1545 (2012).
- Stadtmueller, B. M. & Hill, C. P. Proteasome activators. *Mol. Cell* **41**, 8–19 (2011).
- Li, H. *et al.* Identification of potent and selective non-covalent inhibitors of the *Plasmodium falciparum* proteasome. *J. Am. Chem. Soc.* **136**, 13562–13565 (2014).
- O'Donoghue, A. J. *et al.* Destructin-1 is a collagen-degrading endopeptidase secreted by *Pseudogymnoascus destructans*, the causative agent of white-nose syndrome. *Proc. Natl Acad. Sci. USA* **112**, 7478–7483 (2015).
- O'Donoghue, A. J. *et al.* Global identification of peptidase specificity by multiplex substrate profiling. *Nature Methods* **9**, 1095–1100 (2012).
- Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J. & Gevaert, K. Improved visualization of protein consensus sequences by iceLogo. *Nature Methods* **6**, 786–787 (2009).
- Harris, J. L., Alper, P. B., Li, J., Rechsteiner, M. & Backes, B. J. Substrate specificity of the human proteasome. *Chem. Biol.* **8**, 1131–1141 (2001).
- Kisselev, A. F., van der Linden, W. A. & Overkleeft, H. S. Proteasome inhibitors: an expanding army attacking a unique target. *Chem. Biol.* **19**, 99–115 (2012).
- Bogyo, M. *et al.* Covalent modification of the active site threonine of proteasomal  $\beta$  subunits and the *Escherichia coli* homolog HslV by a new class of inhibitors. *Proc. Natl Acad. Sci. USA* **94**, 6629–6634 (1997).
- da Fonseca, P. C. A. & Morris, E. P. Cryo-EM reveals the conformation of a substrate analogue in the human 20S proteasome core. *Nature Commun.* **6**, 7573–7576 (2015).
- Harshbarger, W., Miller, C., Diedrich, C. & Sacchettini, J. Crystal structure of the human 20S proteasome in complex with carfilzomib. *Structure* **23**, 418–424 (2015).
- Meshnick, S. R. Artemisinin: mechanisms of action, resistance and toxicity. *Int. J. Parasitol.* **32**, 1655–1660 (2002).
- Ashley, E. A. *et al.* Tracking Resistance to Artemisinin Collaboration (TRAC). Spread of artemisinin resistance in *Plasmodium falciparum* malaria. *N. Engl. J. Med.* **371**, 411–423 (2014).
- Stratimer, J. *et al.* Drug resistance. K13-propeller mutations confer artemisinin resistance in *Plasmodium falciparum* clinical isolates. *Science* **347**, 428–431 (2015).
- Dogovski, C. *et al.* Targeting the cell stress response of *Plasmodium falciparum* to overcome artemisinin resistance. *PLoS Biol.* **13**, e1002132 (2015).
- Britton, M. *et al.* Selective inhibitor of proteasome's caspase-like sites sensitizes cells to specific inhibition of chymotrypsin-like sites. *Chem. Biol.* **16**, 1278–1289 (2009).
- Bedford, L., Paine, S., Sheppard, P. W., Mayer, R. J. & Roelofs, J. Assembly, structure, and function of the 26S proteasome. *Trends Cell Biol.* **20**, 391–401 (2010).
- Palmer, J. T., Rasnick, D., Klaus, J. L. & Brömme, D. Vinyl sulfones as mechanism-based cysteine protease inhibitors. *J. Med. Chem.* **38**, 3193–3196 (1995).
- Arastu-Kapur, S. *et al.* Identification of proteases that regulate erythrocyte rupture by the malaria parasite *Plasmodium falciparum*. *Nature Chem. Biol.* **4**, 203–213 (2008).
- Zhou, H. J. *et al.* Design and synthesis of an orally bioavailable and selective peptide epoxyketone proteasome inhibitor (PR-047). *J. Med. Chem.* **52**, 3028–3038 (2009).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was supported by National Institutes of Health grants R01AI078947, R01EB05011 to M.B., and by the Medical Research Council grant MC-UP-1201/5 to P.C.A.d.F. H.L. was supported by an NNS-PhD scholarship from the Agency for Science, Technology and Research (A\*STAR) Singapore. W.A.v.d.L. was supported by a Rubicon fellowship from the Netherlands Organization for Scientific Research (NWO). A.J.O. and C.S.C. were supported by the Program for Breakthrough Biomedical Research (PBBR) and the Sandler Foundation. I.T.F. was supported by American Heart Association grant 14POST20280004. We acknowledge support from the Australian Research Council and the Australian National Health and Medical Research Council. We thank K. Chotivanich for providing PL2 and PL7 parasites. We thank E. Yeh's group for help with *P. falciparum* D10 culture and for use of their equipment. We thank J. Boothroyd for providing the human fibroblast cells. We thank E. Morris and R. Henderson for discussions on image processing, FEI fellows and C. Sava for assisting in the use of the Titan Krios microscope, S. Chen for EM support, and J. Grimmet and T. Darling for computing support.

**Author Contributions** H.L., A.J.O., L.T., C.S.C., P.C.A.d.F. and M.B. designed the experiments and wrote the manuscript. H.L., W.L. and E.Y. performed chemical synthesis and analysis. H.L., S.C.X. and I.T.F. performed the inhibitor studies. P.C.A.d.F. did the electron microscopy and image analysis. All authors discussed the results and commented on the manuscript.

**Author Information** The cryo-EM map and the atomic coordinates of the inhibitor-bound *Plasmodium* 20S proteasome have been deposited in the Electron Microscopy Data Bank (EMDB) and PDB under accession numbers EMD-3231 and 5FMG, respectively. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.B. ([mbogyo@stanford.edu](mailto:mbogyo@stanford.edu)) or P.C.A.d.F. ([pauladf@mrc-lmb.cam.ac.uk](mailto:pauladf@mrc-lmb.cam.ac.uk)).

## METHODS

**Materials for substrate profiling.** Proteasome substrates Suc-LLVY-amc and human 20S proteasome were purchased from Boston Biochem. The *P. falciparum* 20S proteasome was purified as previously reported<sup>11</sup>. The human PA28 $\alpha$  was a gift from Boston Biochem. The peptide library for substrate cleavage profiling was used as described in ref. 15.

**Fluorogenic substrate assay.** Substrate assays were performed in black 96-well Costar plates. For these assays, purified human or *P. falciparum* 20S proteasomes were activated using human PA28 $\alpha$ <sup>13</sup>. Each enzyme was added to a final concentration of 1 nM and the reaction proceeded in the presence of 12.5  $\mu$ M Suc-LLVY-amc for 1 h at room temperature (21 °C). The slope over the linear range of the reaction was determined for each of the enzymes. All experiments were performed at least in triplicate.

**Multiplex substrate profiling of human and *Plasmodium* proteasomes.** The peptide library consisting of 228 tetradecapeptides was designed according to an algorithm previously reported<sup>15</sup>. Peptides were split into 2 pools of 114 for optimal liquid chromatography separation. *P. falciparum* and human 20S proteasomes were assayed at 2 nM, in the presence of 24 nM of PA28 $\alpha$  and 500 nM of each peptide. All assays were performed at room temperature in 20 mM HEPES, pH 7.5, and 0.5 mM EDTA. Aliquots were removed and acid quenched to pH 3.0 or less with formic acid (4% final) after incubation for 15, 60, and 240 min. A control sample lacking enzyme was also prepared under identical conditions to account for non-enzymatic degradation of the substrates. Each sample was de-salted with C18 tips (Rainin), and ~130 ng of total peptide was injected for mass spectrometry analysis. Peptide separation was performed using an EZ-Spray C18 column (Thermo, ES800, PepMap, 3  $\mu$ m bead size, 75  $\mu$ m  $\times$  15 cm). Reverse-phase chromatography was performed at a flow rate of 600 nl min<sup>-1</sup> for loading and 300 nl min<sup>-1</sup> for using a nanoACQUITY (Waters) ultra-performance liquid chromatography instrument. A 60 min linear gradient from 2% to 30% acetonitrile in 0.1% formic acid was used, and mass spectrometry was performed on a LTQ Orbitrap XL. Survey scans were recorded over a 350–1500 *m/z* range, and peptide fragmentation was induced with collision-induced dissociation for tandem mass spectrometry. Collision-induced dissociation was performed on the six most intense precursor ions, with a minimum of 1,000 counts, using an isolation width of 2.0 atomic mass units, with 35% normalized collision energy. Data were analysed as previously reported using the in-house PAVA software to generate centroid peak lists, Protein Prospector software version 5.10.15 for data searches, and iceLogo software for visualization of amino-acid frequency surrounding the cleavage sites<sup>15</sup>. The Z-score was calculated using the standard formula and indicates deviation from the mean frequency of a specific amino acid on a specific position in a reference set of sequences.

***P. falciparum* replication assays.** *P. falciparum* D10 cultures were maintained, synchronized, and lysed as previously described<sup>11</sup>. *P. falciparum* D10 was obtained originally from the Malaria Research and Reference Reagent Resource Center (MR4). Parasite cultures were grown in human erythrocytes (no blood type discrimination) purchased from the Stanford Blood Center (approved by Stanford University). Ring stage *P. falciparum* culture at 2% parasitaemia and 0.5% hematocrit was added to 96-well plates spotted with compounds. The *P. falciparum* culture was incubated with each compound for 1 h and 72 h. After compound treatment for 1 h, the inhibitor was removed and the culture was washed three times in media, before fresh media was added for a further 71 h of growth. After incubation, the culture was fixed in a final concentration of 1% paraformaldehyde (in PBS) for 30 min at room temperature. The nuclei stain YOYO-1 was then added to a final concentration of 50 nM and incubated overnight at 4 °C. Parasite replication was monitored by observation of a YOYO-1-positive population (infected) and YOYO-1-negative population (uninfected) using a BD Accuri C6 automated flow cytometer. Values in Extended Data Fig. 1e are the average  $\pm$  s.d.; *n* = 6 parasite cultures from two independent experiment of triplicates.

**Cell viability assay in mammalian cells.** HFFs used in all experiments were non-transformed primary cells (obtained from J. Boothroyd) derived from human foreskin tissues and tested for mycoplasma contamination. HFFs were plated at 2,500 cells (non-confluent) per well <24 h before addition of the compound. Compounds were diluted for dose–response concentrations in media, and added to the cells for 72 h. Cell viability was measured using the CellTiter-Blue Assay (Promega) as per the manufacturer's instructions. Values in Extended Data Fig. 1e are the average  $\pm$  s.d.; *n* = 9 cell cultures from three independent experiments of triplicates for HFF treatment, except for 1 h WLW-vs, 1 h LLW-vs, and 72 h LLL-vs, where *n* = 6 cell cultures from two independent experiment of triplicates.

**Activity-based probe labelling of mammalian and *P. falciparum* proteasomes.** BMV037 was used at a final concentration of 10  $\mu$ M to label *P. falciparum* proteasome activity, and MV151 was used at a final concentration of 2  $\mu$ M to label mammalian proteasome activity. Purified proteasomes (human or *P. falciparum*, 10 nM) were incubated with each inhibitor for 1 h at 37 °C before addition of

BMV037 or MV151 for a further 2 h at 37 °C. Samples were denatured by addition of SDS sample buffer, boiled briefly and run on a 12% SDS PAGE. Gels were scanned on the Cy5 channel (for BMV037) or the Cy3 channel (for MV151) on a Typhoon Scanner (GE Healthcare). Quantification of the intensity of the labelled proteins was done using Image J. A small amount of activation in the  $\beta$ 1-subunit was observed in the human proteasome after inhibitor treatment.

**Assessment of *P. falciparum* proteasome activity in intact inhibitor treated parasites.** *P. falciparum* was cultured to around 10% parasitaemia to ensure sufficient parasites were available for proteasome labelling. *P. falciparum* culture was first treated for the indicated amount of time, and spun down at 3,200 r.p.m (962g) to remove inhibitor. Each culture was then washed once in fresh media and *P. falciparum* lysate obtained as previously described<sup>11</sup>. Proteasome labelling of the parasite lysate was performed as described above. To assess proteasome inhibition, the intensity of the proteasome labelling was quantified by Image J and the amount of proteasome inhibition was taken as a percentage of the dimethylsulfoxide (DMSO)-treated control.

**Assessment of WLL-vs activity and DHA synergism against ART-sensitive and -resistant *P. falciparum*.** Sensitivity of PL2 and PL7 *P. falciparum* strains (at 2 h after invasion) to a 1-h pulsed treatment with WLW-vs was determined as described previously<sup>25</sup>. Interactions between DHA and WLW-vs against the K13 mutant strain, PL7, exposed to a 3-h pulsed treatment was determined at 4 h after invasion as described previously<sup>25</sup>.

**In vivo assessment of WLL-vs.** All mouse experiments were approved by the Stanford Administration Panel on Laboratory Animal Care and strictly followed their specific guidelines. Sample size was determined on the basis of the minimum number of animals required for good data distribution and statistics. We ensured all animals in vehicle- and compound-treated groups had similar body weight before treatment. No inclusion or exclusion criteria were used as animals were randomly chosen for each group. No blinding was performed between the groups. For each drug assay, Balb/c female mice (~20 g, 6–8 weeks old) were infected by intraperitoneal injection with  $1 \times 10^6$  *P. chabaudi* parasites isolated from an infected mouse on day 0. For the experiment presented in Fig. 4, the infected mice were dosed 3 days after infection with a single bolus tail vein injection either of vehicle (*n* = 6) comprising 10% DMSO, 10% (2-hydroxypropyl)- $\beta$ -cyclodextrin in sterile PBS, or of WLL-vs (*n* = 6) formulated in the vehicle. For the experiment presented in Extended Data Fig. 4c, the infected mice were dosed 2 days after infection with a single bolus intraperitoneal injection either of vehicle (45% polyethylene glycol (relative molecular mass 400), 35% propylene glycol, 10% ethanol, 10% DMSO and 10% (w/v) 2-hydroxypropyl- $\beta$ -cyclodextrin), or WLL-vs at 40 mg kg<sup>-1</sup>, 60 mg kg<sup>-1</sup>, and 80 mg kg<sup>-1</sup> body weight formulated in the vehicle. Both of the treatments were administered at night (21:00). Parasitemia was monitored daily by Geimsa-stained thin blood smears obtained from the tail vein and quantified by light microscope counting, where a minimum of five fields, each containing at least 400 red blood cells, were counted. The health of the mice was also monitored daily by measuring their body weight and by assessment of their appearance and activity.

**Assessment of *P. falciparum* cysteine protease activities.** A mixed stage culture of red blood cells infected with *P. falciparum* D10 was treated with a range of concentrations of WLL-vs for 1 h at 37 °C. BODIPY-TMR-DCG04 (2  $\mu$ M) was then added to the culture for a further 1 h for live cell labelling of the activities of the parasitic cysteine proteases. JPM-OEt<sup>29</sup> (100  $\mu$ M) was added as a positive control. After probe labelling, the parasite pellet was isolated by saponin lysis and kept at –80 °C. For further analysis, the samples were thawed at 4 °C and lysed directly in sample buffer and loaded onto a 15% SDS–polyacrylamide gel electrophoresis. Fluorescence of the probe was detected using the Typhoon as described above.

**Giemsa stain of WLL-vs treated *P. falciparum* culture.** *P. falciparum* culture at mid-ring (9–12 h after invasion) was treated with DMSO or WLL-vs for 1 h at 37 °C and the inhibitor was removed. The culture was washed in media and left to grow in fresh media for a further 23 h. Culture was then spun down and the red blood cell pellet used to make a thin blood smear on glass slides, fixed with methanol and stained with Giemsa stain for light microscopy.

**Chemical synthesis.** The detailed procedures for chemical synthesis are presented in Supplementary Information.

**Cryo-EM data acquisition.** A sample of *P. falciparum* 20S proteasome core was diluted in 50 mM Tris-HCl, pH 7.5, 5 mM MgCl<sub>2</sub>, and 1 mM dithiothreitol, to a concentration of 0.14  $\mu$ M (0.1 mg ml<sup>-1</sup>), and incubated in the presence of 10  $\mu$ M WLW-vs for 1 h at 37 °C. The sample was applied onto Quantifoil 1.2/1.3 electron microscope grids that were coated with a thin layer of carbon, freshly floated from mica, and glow-discharged in air. The grids were flash frozen by plunging into liquid ethane using a Vitrobot (FEI), operated at 22 °C and 95% humidity, with a 2.5 s blot time. The grids were transferred into a FEI Titan Krios microscope, operated at an acceleration voltage of 300 keV and a nominal magnification of  $\times 75,000$  (yielding a calibrated sampling of 1.04 Å per pixel at the image level).

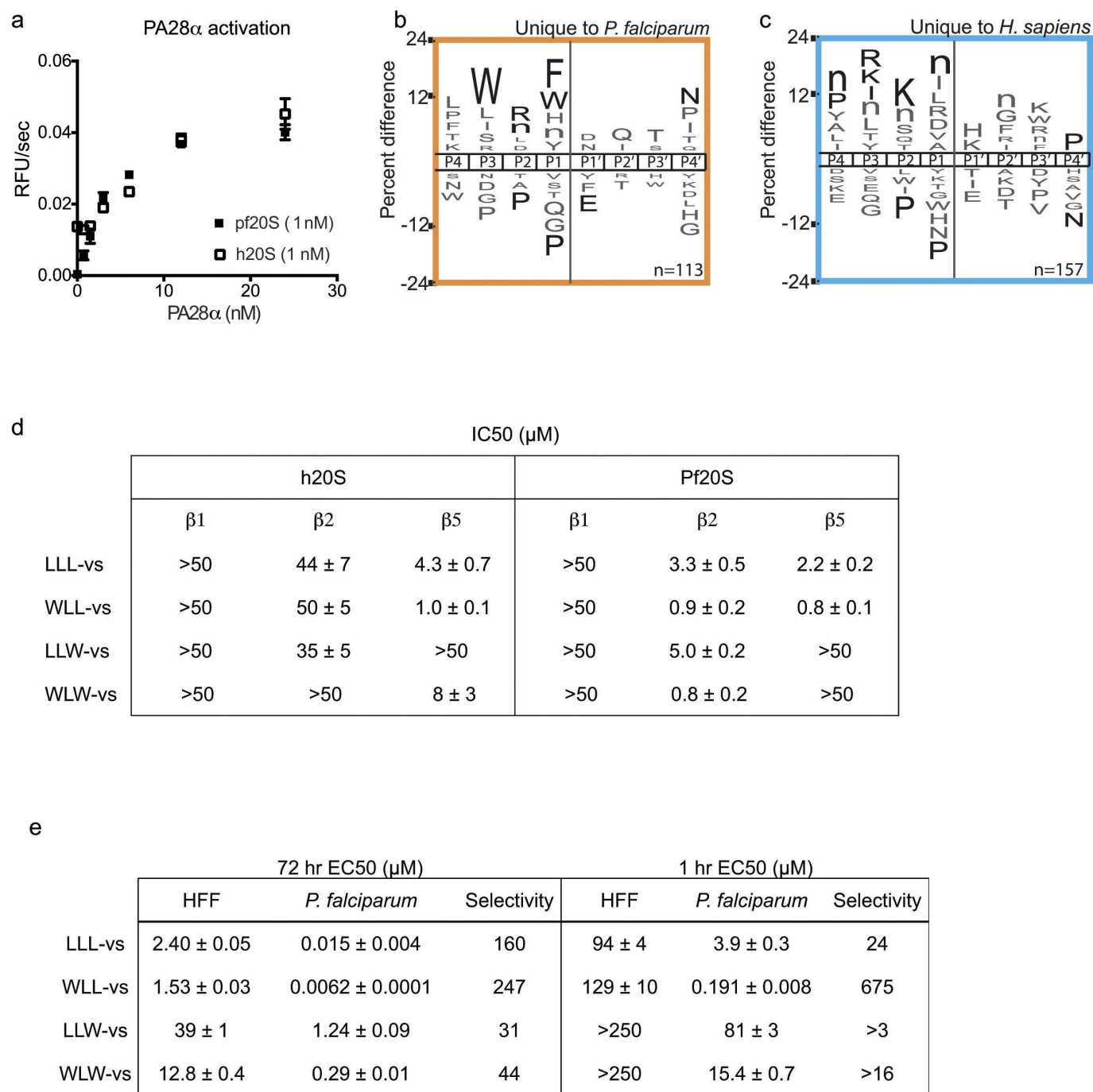


Images were recorded (from a single cryo-EM grid and during a single cryo-EM session) on a Falcon II direct electron detector, using EPU software, at a range of 1.70–3.16  $\mu\text{m}$  underfocus and 1 s exposure time (Extended Data Fig. 3a). For each exposure 17 individual frames were captured at a rate of 0.056 s per frame, each corresponding to an electron dose of 2.8 electrons per square ångström. For each exposure, the image corresponding to the sum of all frames was used for screening for optimal ice thickness and image stability, as assessed by the isotropic recovery of Thon rings in their power spectra. From the 2,603 images recorded, 1,816 were selected for further processing.

**Image processing.** We used the same strategy for image analysis as for the recent determination of the structure of an inhibitor-bound human 20S proteasome core<sup>20</sup>. For each exposure selected for image analysis, the sum of all frames captured was used to determine the defocus and astigmatism parameters, using the Tigris program findctf, and for particle selection, performed automatically in Relion<sup>31</sup>. This automated particle selection was carefully supervised with the manual removal of false positives and addition of false negatives. Further processing was performed using the sum of frames 3–10 of each exposure, effectively acquired within 0.45 s and corresponding to a total accumulated electron dose of 28 electrons per square ångström. This frame selection aimed at an adequate signal-to-noise ratio, for accurate determination of particle orientation, and the preservation of high-resolution information, as previously discussed<sup>20</sup>. The phases of the resulting images were corrected for the contrast transfer function by phase flipping with the Tigris program flipctf, using the image defocus and astigmatism parameters determined as described above. Selected particle images were then extracted into 256 pixel  $\times$  256 pixel boxes, resulting in a data set of 97,720 particles. The single-particle analysis refinement routine was performed using C2 symmetry and used the cryo-EM structure of the human 20S proteasome core (Electron Microscopy Data Bank accession code EMD-2981) as starting reference. This routine consisted of rounds of alignment and Euler angle assignment by projection matching, using the program AP SH of the Spider software package<sup>32</sup>, 3D reconstruction, using icr3d (which includes correction for the amplitude component of the contrast transfer function), and forward projection, using icr3dpro. The icr3d and icr3dpro programs are implemented in Tigris, a software package publicly available at <http://sourceforge.net>. A resolution limited approach was used to avoid noise over-fitting during refinement, where the 3D map obtained in each refinement cycle was carefully inspected, masked, and Fourier low-pass filtered to a conservative threshold in preparation of the reference for the next cycle. The surface representations of the final map as determined by the icr3d software (Fig. 2a, b), without further sharpening, Fourier filtering, or masking, were created using UCSF Chimera<sup>33</sup> and colour coded according to local resolution as determined by ResMap<sup>34</sup>.

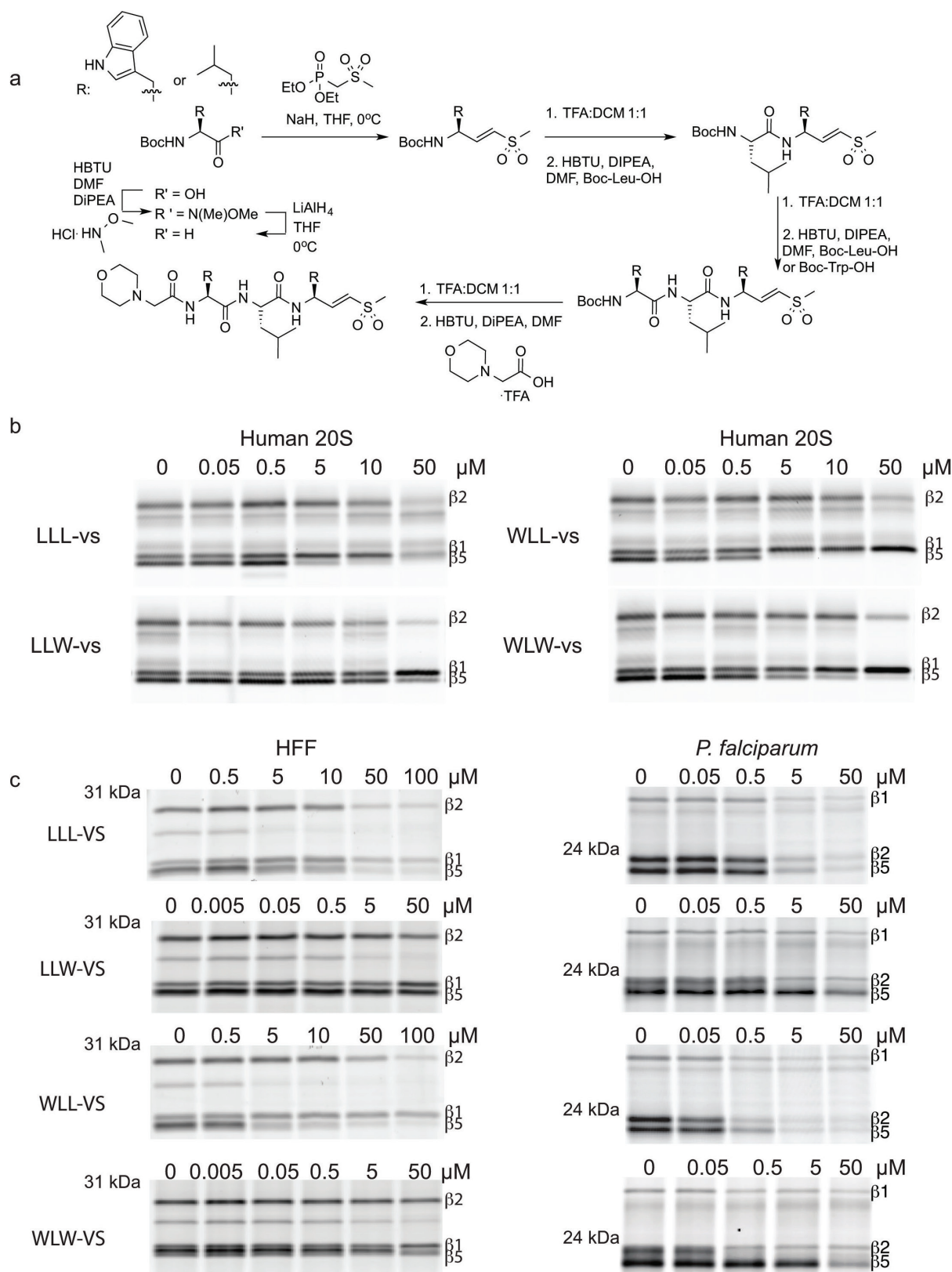
**Molecular modelling.** Molecular models for each of the *P. falciparum* proteasome 20S core subunits were originally obtained using the Phyre2 Protein Homology Recognition Engine<sup>35</sup>, with the sequences deposited in the PlasmoDB (*Plasmodium* Genomics Resource). A preliminary fitting of these models was performed using molecular dynamics flexible fitting<sup>36</sup>, which was followed by real-space refinement in Coot<sup>37</sup> and Phenix<sup>38</sup>. Protein regions that were not well recovered in our map were not modelled. The final model was checked for geometry, close contacts, and bond parameters using MolProbity<sup>39</sup>. The statistics obtained (Extended Data Fig. 3c) compared favourably with those for models determined by X-ray crystallography at similar resolutions, indicating that there was no over-fitting during model building. The resolution of the cryo-EM map can therefore be assessed by Fourier shell correlation against a density map generated from these coordinates, yielding a value of about 3.6 Å (Extended Data Fig. 3d). The mesh representations of the final map showing the fitted coordinates of the *P. falciparum* 20S proteasome core (Fig. 2c–g) were created using the PyMOL Molecular Graphics System. For these representations, the map was sharpened with a *B*-factor of –50 and Fourier low-pass filtered to 3.6 Å.

31. Scheres, S. H. W. Semi-automated selection of cryo-EM particles in RELION-1.3. *J. Struct. Biol.* **189**, 114–122 (2015).
32. Frank, J. et al. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
33. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
34. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nature Methods* **11**, 63–65 (2014).
35. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**, 845–858 (2015).
36. Trabuco, L. G., Villa, E., Mitra, K., Frank, J. & Schulten, K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16**, 673–683 (2008).
37. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
38. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352–367 (2012).
39. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
40. Verdoes, M. et al. A fluorescent broad-spectrum proteasome inhibitor for labeling proteasomes in vitro and in vivo. *Chem. Biol.* **13**, 1217–1226 (2006).
41. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).



**Extended Data Figure 1 | Activity, substrate specificity and inhibition of human and *P. falciparum* 20S proteasome.** **a–c**, Substrate cleavage profile of activated human and *P. falciparum* 20S proteasome. **a**, Activation of the human and *P. falciparum* 20S proteasome by human PA28 $\alpha$ . Activity was determined by cleavage of the fluorogenic substrate Suc-LLVY-amc. Error bars, s.d.  $n = 3$  purified proteasome in technical replicates. **b**, **c**, iceLogos of cleavage sequences that are uniquely processed either by the *P. falciparum* (**b**) or by the human proteasome (**c**). Amino acids that are most and least favoured at each position are shown above and below the axis, respectively. Lower-case 'n', norleucine; amino acids in black text are statistically significant ( $P < 0.05$ , unpaired two-tailed Student's  $t$ -test). **d**, **e**, Inhibition potencies of the vinyl sulfone inhibitors. **d**, Table of IC<sub>50</sub> values for each inhibitor in 1-h treated *P. falciparum* and human

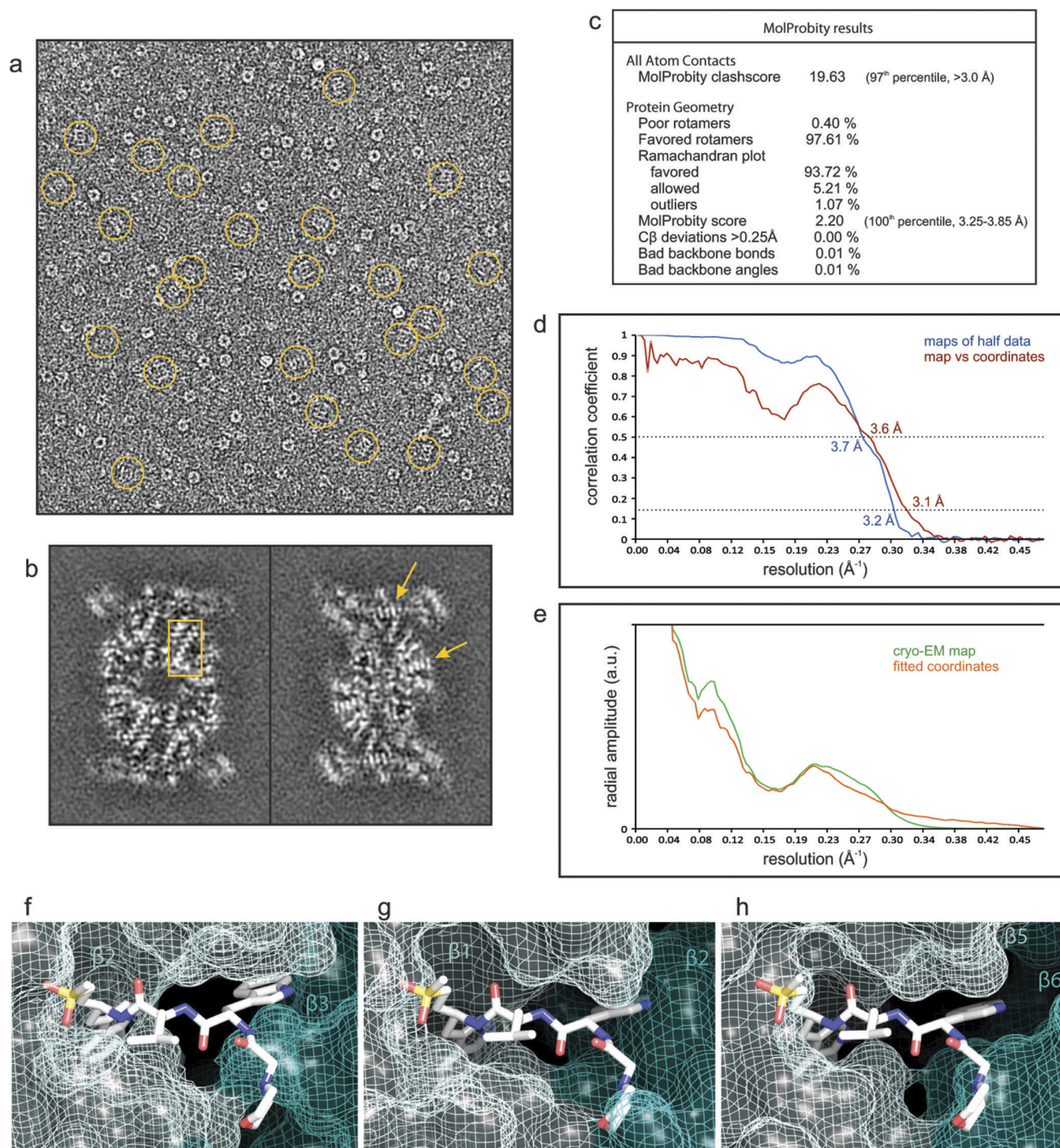
20S proteasome. IC<sub>50</sub> values are determined from three independent experiments of inhibitor pretreatment followed by activity labelling of the 20S proteasome ( $n = 3$  purified proteasome). Gels in Fig. 1f and Extended Data Fig. 2b were quantified to calculate the IC<sub>50</sub> values (for gel source data and replicates, see Supplementary Fig. 1a, b). Data are mean  $\pm$  s.d. **e**, Table of EC<sub>50</sub> values for each of the inhibitors in 1 h and 72 h treatment of *P. falciparum* at ring stage or non-confluent HFFs. Data are mean  $\pm$  s.d.;  $n = 6$  parasite cultures from two independent experiment of triplicates for *P. falciparum* treatments;  $n = 9$  cell cultures from three independent experiments of triplicates for HFF treatment, except for 1 h WLW-vs, 1 h LLW-vs, and 72 h LLL-vs, where  $n = 6$  cell cultures from two independent experiments of triplicates.



**Extended Data Figure 2 | Proteasome inhibitors preferentially inhibit  $\beta 2$  of the *P. falciparum* proteasome.** **a**, Vinyl sulfone inhibitors are synthesized from the Boc-protected amino acid by first generating the Weinreb amide, followed by the Horner–Wadsworth–Emmons reaction and standard peptide coupling. **b**, Purified human 20S proteasome was pre-treated for 1 h at 37 °C with each inhibitor followed by addition of activity-based probe MV151 (ref. 40) to assess for human proteasome

activities (for gel source data, see Supplementary Fig. 1b). **c**, HFF or *P. falciparum* culture was treated for 1 h at 37 °C with each inhibitor, followed by compound washout and post-lysis activity-based probe labelling. Gel shown for WLL-vs in *P. falciparum* is derived from Fig. 4c at the indicated concentrations to allow for direct comparison with other compounds (for gel source data, see Supplementary Fig. 1e).



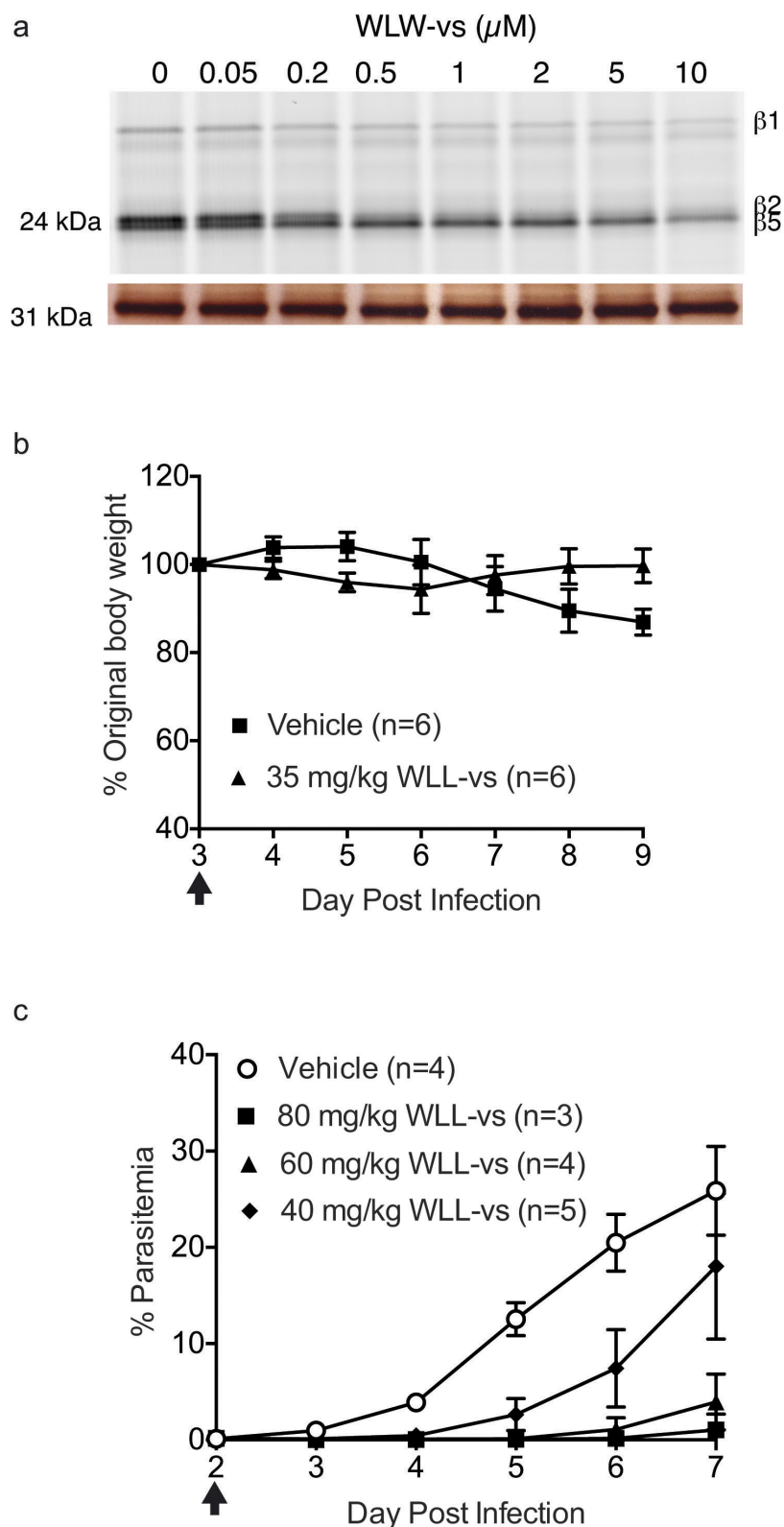


### Extended Data Figure 3 | Evaluation of the final cryo-EM map and molecular model of the *P. falciparum* 20S proteasome core.

**a–e**, Evaluation of the single-particle analysis of the *P. falciparum* 20S proteasome core bound to the inhibitor WLW-vs. **a**, Cryo-EM image of the sample analysed, with molecular images of side views of the complex (normal to its long axis) indicated by rings. The image greyscale was inverted to show the protein densities in white. **b**, Individual sections of the 3D map, as determined by the 3D reconstruction algorithm (without further sharpening, masking, or Fourier filtering), are represented as grey scale. These sections are 1 Å thick and reveal the quality of the reconstruction, as the protein densities are clearly resolved against a very smooth background, with regions showing the pattern of  $\alpha$ -helices (box) and the clear separation of sheet-forming  $\beta$ -strands (arrows) indicated. **c**, Evaluation of the model of the *P. falciparum* 20S proteasome core using MolProbity<sup>39</sup>. **d**, Resolution estimate of the cryo-EM map by Fourier shell correlation. The curves correspond to the correlation obtained against the protein model (red) and the correlation between maps determined from

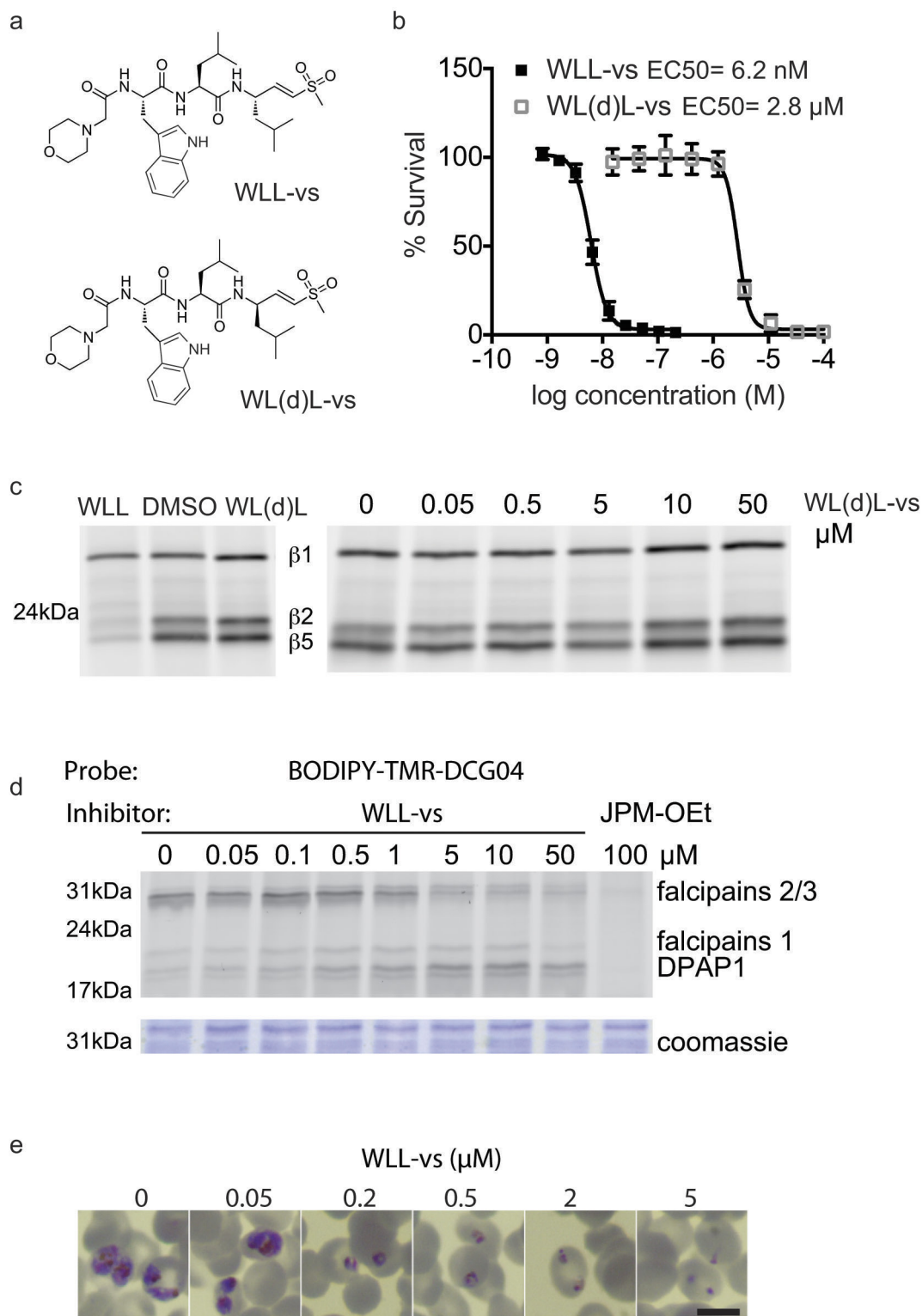
two halves of the data (blue). The resolution was estimated from the curve against the model where the 0.5 correlation coefficient criterion<sup>41</sup> yields an estimate of 3.6 Å. The correlation coefficient can be seen to fall to a local minimum at ~6 Å and then recover at higher resolutions for both Fourier shell correlation curves. This behaviour is consistent with the rotationally averaged amplitude spectra of both the cryo-EM map and the coordinates (e). This region of the amplitude spectra contains reduced structural information, typical of protein scattering, indicating that these effects in the Fourier shell correlation curves arise from a genuine local reduction in the signal:noise ratio. **f–h**, Accessibility of the human 20S proteasome active sites to the inhibitor WLW-vs, using the protein model of the human proteasome core complex bound to an LLL-vs inhibitor<sup>20</sup>. Protein coordinates of the human proteasome 20S core (PDB 5A0Q)  $\beta$ 2 (**f**),  $\beta$ 1 (**g**), and  $\beta$ 5 (**h**) active sites were aligned to the coordinates of the *P. falciparum* proteasome  $\beta$ 2-subunit bound to the WLW-vs inhibitor. The model of the human 20S proteasome active sites is represented as van der Waals surfaces with the superimposed WLW-vs inhibitor shown as sticks.





**Extended Data Figure 4 | Intact cell treatment and *in vivo* treatment of vinyl sulfone inhibitors.** **a**, WLW-vs was incubated in early trophozoite *P. falciparum* culture for 3 h, washed out, and the parasite lysate was incubated with probe BMV037. Top gel, the fluorescent scan; bottom gel, the silver stain. For gel source data, see Supplementary Fig. 1f. **b**, Body weight of WLL-vs- and vehicle-treated Balb/c mice after compound treatment by tail vein injection (Fig. 4e), expressed as a percentage of the original body weight on day 3 before compound treatment. Body weight of vehicle-treated mice decreased after day 6 of infection as part of the response to the natural resolution of the *P. chabaudi* infection. Treatment

day is indicated by arrow;  $n = 6$  mice for each group; error bars, s.d. **c**, Balb/c female mice infected with  $1 \times 10^6$  *P. chabaudi* parasites from passage host on day 0 were treated with a single bolus dose of vehicle (45% polyethylene glycol (relative molecular mass 400), 35% propylene glycol, 10% ethanol, 10% DMSO, and 10% (w/v) 2-hydroxypropyl- $\beta$ -cyclodextrin;  $n = 4$  mice) or WLL-vs at  $40 \text{ mg kg}^{-1}$  ( $n = 5$  mice),  $60 \text{ mg kg}^{-1}$  ( $n = 4$  mice), and  $80 \text{ mg kg}^{-1}$  ( $n = 3$  mice) formulated in the vehicle. Treatment was performed on day 2 after infection as indicated by the arrow and administered by intraperitoneal injection. Parasitaemia was monitored daily by Giemsa stain of thin blood smears. Error bars, s.d.



#### Extended Data Figure 5 | Assessing off-target activities of WLL-vs.

**a**, Structures of WLL-vs and its diastereomer WL(d)L-vs. **b**, Dose-response curves of WLL-vs and WL(d)L-vs after 72 h treatment in *P. falciparum*. Error bars, s.d. ( $n = 6$  parasite cultures for WLL-vs from triplicates of two independent experiments, and  $n = 8$  parasite cultures for WL(d)L-vs over three independent experiments). **c**, Purified *P. falciparum* 20S proteasome was treated for 1 h at 37 °C with 10 μM of WLL-vs and WL(d)L-vs (left) or a range of concentrations of WL(d)L-vs. Residual activity was assessed by probe BMV037 (for gel source data, see Supplementary Fig. 1g).

**d**, A mixed-stage culture of *P. falciparum* was treated for 1 h with WLL-vs at 37 °C, followed by BODIPY-TMR-DCG04 for a further 1 h. Samples were directly loaded onto SDS-polyacrylamide gel electrophoresis for analysis. JPM-OEt (100 μM) was included as positive control. The fluorescent scan is shown at the top and the Coomassie stain is shown at the bottom. For gel source data, see Supplementary Fig. 1h. **e**, Geimsa stain of 1 h treated *P. falciparum* ring 24 h after inhibitor was added. Scale bar, 600 μm.

# Backbone NMR reveals allosteric signal transduction networks in the $\beta_1$ -adrenergic receptor

Shin Isogai<sup>1</sup>, Xavier Deupi<sup>2</sup>, Christian Opitz<sup>1</sup>, Franziska M. Heydenreich<sup>2</sup>, Ching-Ju Tsai<sup>2</sup>, Florian Brueckner<sup>2</sup>, Gebhard F. X. Schertler<sup>2,3</sup>, Dmitry B. Vepintsev<sup>2,3</sup> & Stephan Grzesiek<sup>1</sup>

G protein-coupled receptors (GPCRs) are physiologically important transmembrane signalling proteins that trigger intracellular responses upon binding of extracellular ligands. Despite recent breakthroughs in GPCR crystallography<sup>1–3</sup>, the details of ligand-induced signal transduction are not well understood owing to missing dynamical information. In principle, such information can be provided by NMR<sup>4</sup>, but so far only limited data of functional relevance on few side-chain sites of eukaryotic GPCRs have been obtained<sup>5–9</sup>. Here we show that receptor motions can be followed at virtually any backbone site in a thermostabilized mutant of the turkey  $\beta_1$ -adrenergic receptor ( $\beta_1$ AR)<sup>10–12</sup>. Labelling with [<sup>15</sup>N]valine in a eukaryotic expression system provides over twenty resolved resonances that report on structure and dynamics in six ligand complexes and the apo form. The response to the various ligands is heterogeneous in the vicinity of the binding pocket, but gets transformed into a homogeneous readout at the intracellular side of helix 5 (TM5), which correlates linearly with ligand efficacy for the G protein pathway. The effect of several pertinent, thermostabilizing point mutations was assessed by reverting them to the native sequence. Whereas the response to ligands remains largely unchanged, binding of the G protein mimetic nanobody NB80 and G protein activation are only observed when two conserved tyrosines (Y227 and Y343) are restored. Binding of NB80 leads to very strong spectral changes throughout the receptor, including the extracellular ligand entrance pocket. This indicates that even the fully thermostabilized receptor undergoes activating motions in TM5, but that the fully active state is only reached in presence of Y227 and Y343 by stabilization with a G protein-like partner. The combined analysis of chemical shift changes from the point mutations and ligand responses identifies crucial connections in the allosteric activation pathway, and presents a general experimental method to delineate signal transmission networks at high resolution in GPCRs.

A thermostabilized, detergent-resistant mutant of  $\beta_1$ AR<sup>10–12</sup> (TS- $\beta_1$ AR; see Methods) was selectively labelled with [<sup>15</sup>N]valine and produced in insect cells without further chemical modifications. Its 28 valine residues are homogeneously distributed across the receptor (Extended Data Fig. 1) at locations suitable to sense ligand binding and receptor activation. Although resonances of main chain atoms are considerably more difficult to observe than those of mobile side chains of surface residues, they are expected to be better reporters of functional, long-range backbone motions. We succeeded to obtain well-resolved TROSY (transverse relaxation-optimized spectroscopy) spectra of the valine <sup>1</sup>H–<sup>15</sup>N backbone resonances of detergent-solubilized TS- $\beta_1$ AR in its apo form and in complexes with six ligands (Extended Data Fig. 1) ranging in their efficacy from antagonists to agonists (Extended Data Table 1). Despite the absence of deuteration and very short T<sub>2</sub> relaxation times (~4 ms for <sup>1</sup>H<sup>N</sup>), 26 valine resonances could be observed with sufficient sensitivity and resolution.

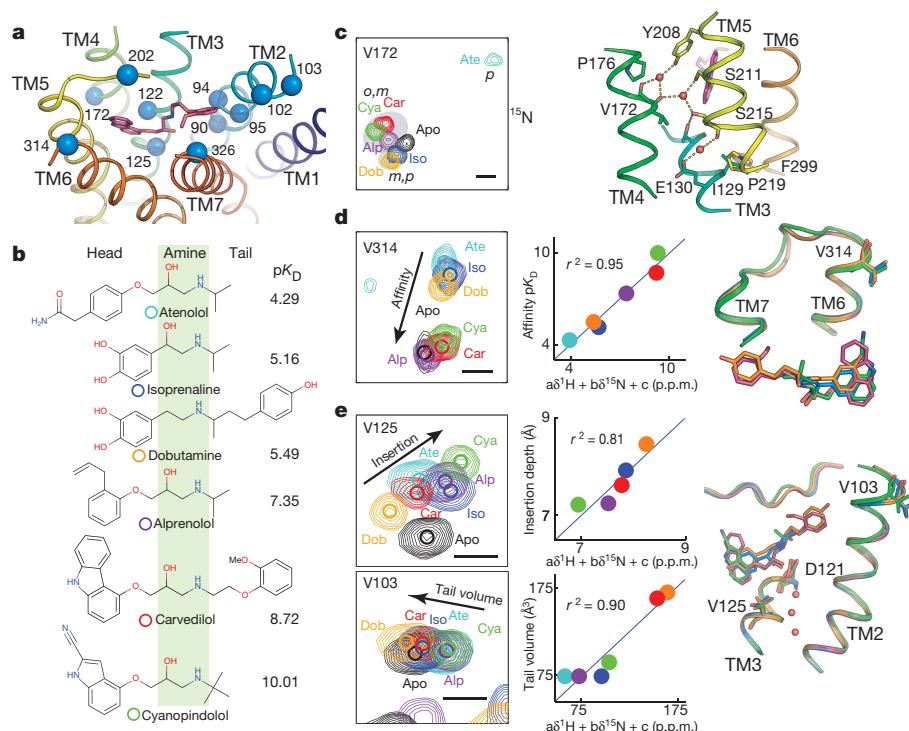
Distinct and reversible chemical shift changes were detected for many valines after ligand exchange. 16 valines were assigned unambiguously and 5 tentatively using spectra from 18 point mutants, as well as further spectral and structural information (Extended Data Table 2).

Many valine residues in the vicinity of the ligand binding pocket could be assigned, showing chemical shift changes that report on the ligand functional groups (Fig. 1). Remarkably, residue V172(4.56) (the number in parenthesis corresponds to the Ballesteros–Weinstein numbering system<sup>13</sup>), which is located close to the ligand aromatic head group, exhibits an unusual <sup>15</sup>N chemical shift of ~105–110 p.p.m. (Extended Data Figs 1, 2 and Fig. 1c). This anomaly seems caused by a distorted backbone geometry, which is presumably conserved among adrenergic receptors and results from a missing hydrogen bond to the proline at position 176(4.60) (Extended Data Fig. 2). Instead, the carbonyl of V172(4.56) participates in a water-mediated hydrogen bond network, which connects the ligand binding site, TM3, TM4, TM5, and TM6<sup>12,14</sup>. Seemingly as a result of these interactions, the V172(4.56) <sup>1</sup>H–<sup>15</sup>N resonances cluster according to the substitution patterns of the ligand head group (Fig. 1c): one cluster is observed for the partial agonists/antagonists cyanopindolol, alprenolol and carvedilol, which have larger head groups with *ortho*- and/or *meta*-substitutions; a second cluster is observed for the agonists isoprenaline and dobutamine, which bear a *meta*- and *para*-substituted catechol ring. We attribute the distinct chemical shifts for isoprenaline or dobutamine to the loss of a coordinated water caused by specific hydrogen bond interactions between their catechol moieties and the side chain of S215(5.461) (Extended Data Fig. 2d).

Remarkably, the resonance positions of V172(4.56) in complex with the antagonist atenolol strongly differ from the already described complexes and the apo form: considerable <sup>1</sup>H<sup>N</sup> (>0.4 p.p.m.) and <sup>15</sup>N (>4 p.p.m.) upfield shifts indicate, respectively, a further weakening of the main chain hydrogen bond V172(4.56)H<sup>N</sup>...I168(4.52)O and a stronger kink of the backbone. This rearrangement seems caused by the insertion of the *para*-acetamide group of the ligand head between residues S215(5.461) and V172(4.56) (Extended Data Fig. 2d). This will lead to a substantial disruption of the TM3–TM4–TM5 interface, thereby precluding receptor activation, in agreement with atenolol's inverse agonist pharmacology. Thus, the amide chemical shifts of V172(4.56) constitute a very sensitive readout for the state of this water-mediated, inter-helical activation switch.

The <sup>1</sup>H–<sup>15</sup>N chemical shifts of further residues in the vicinity of the binding pocket report on additional characteristics of the ligands. V314(6.59) and V202(ECL2) are located at the extracellular surface of the receptor in a “vestibule” next to the entry/exit pathway of the orthosteric binding site<sup>15,16</sup>. The resonances of these residues are either severely broadened or undetectable in the absence of ligands, whereas they are observable in the presence of ligands (Fig. 1d and Extended Data Fig. 3). A line shape analysis for V314(6.59) (Extended Data Fig. 4)

<sup>1</sup>Focal Area Structural Biology and Biophysics, Biozentrum, University of Basel, CH-4056 Basel, Switzerland. <sup>2</sup>Paul Scherrer Institute, CH-5232 Villigen PSI, Switzerland. <sup>3</sup>Department of Biology, ETH Zurich, CH-8093 Zurich, Switzerland.



**Figure 1 | Ligand-induced  $^1\text{H}$ - $^{15}\text{N}$  chemical shift changes in the vicinity of the ligand binding pocket of  $\beta_1\text{AR}$ .** **a**, Partial view of the  $\beta_1\text{AR}$ -carvedilol crystal structure (4AMJ) showing valine residues (blue spheres) in the vicinity ( $<8.5$  Å) of the ligand (magenta sticks) binding site. **b**, Chemical structures of the  $\beta_1\text{AR}$  ligands used in this study. Ligand affinities derived from whole-cell binding assays on the thermostabilized  $\beta_36$ -m23  $\beta_1\text{AR}$  construct<sup>17</sup> are indicated as  $pK_D$  values. Similar  $pK$  values were measured for the TS- $\beta_1\text{AR}$  construct (Extended Data Table 1). **c**, Left, ligand-induced response of V172(4.56)  $^1\text{H}$ - $^{15}\text{N}$  resonances. The black bar represents a scale of 0.1 p.p.m. ( $^1\text{H}$ ) and 1 p.p.m. ( $^{15}\text{N}$ ). The labels *o*, *m*, *p* indicate ligands with *ortho*, *meta*, and *para* substitutions at the head group, respectively. **c**, Right, partial view of the  $\beta_1\text{AR}$ -carvedilol structure (4AMJ) showing the interaction network connecting V172(4.56) to S215(5.461), P219(5.50), I129(3.40), and F299(6.44). **d**, Left, representation as **c**, left, for the  $^1\text{H}$ - $^{15}\text{N}$  resonances of V314(6.59). Centres of resonances are indicated

by circles. **d**, Middle, correlation of a best-fit linear combination of the V314(6.59) chemical shifts ( $-49.3 \delta^1\text{H} + 2.02 \delta^{15}\text{N} + 133$ ) to the ligand affinity  $pK_D$  (Extended Data Table 1). **d**, Right, partial view of the carvedilol (red, 4AMJ), dobutamine (orange, 2Y01), cyanopindolol (green, 2VT4 B), and isoprenaline (blue, 2Y03) complex structures showing the ligand-induced movement of V314(6.59). **e**, Left, representation as **d**, left, for the  $^1\text{H}$ - $^{15}\text{N}$  resonances of V125(3.36) and V103(2.65). **e**, Middle, correlations of best-fit linear combinations of chemical shifts for V125(3.36) ( $-0.402 \delta^1\text{H} + 1.17 \delta^{15}\text{N} - 132$ ) and V103(2.65) ( $-336 \delta^1\text{H} - 634 \delta^{15}\text{N} + 7.64 \cdot 10^4$ ) to the ligand insertion depths and tail volumes (Extended Data Table 1), respectively. **e**, Right, partial view of the carvedilol, dobutamine, cyanopindolol, and isoprenaline crystal structures (representation as **d**, right) showing the ligand-induced movement of V125(3.36), D121(3.32), and V103(2.65).

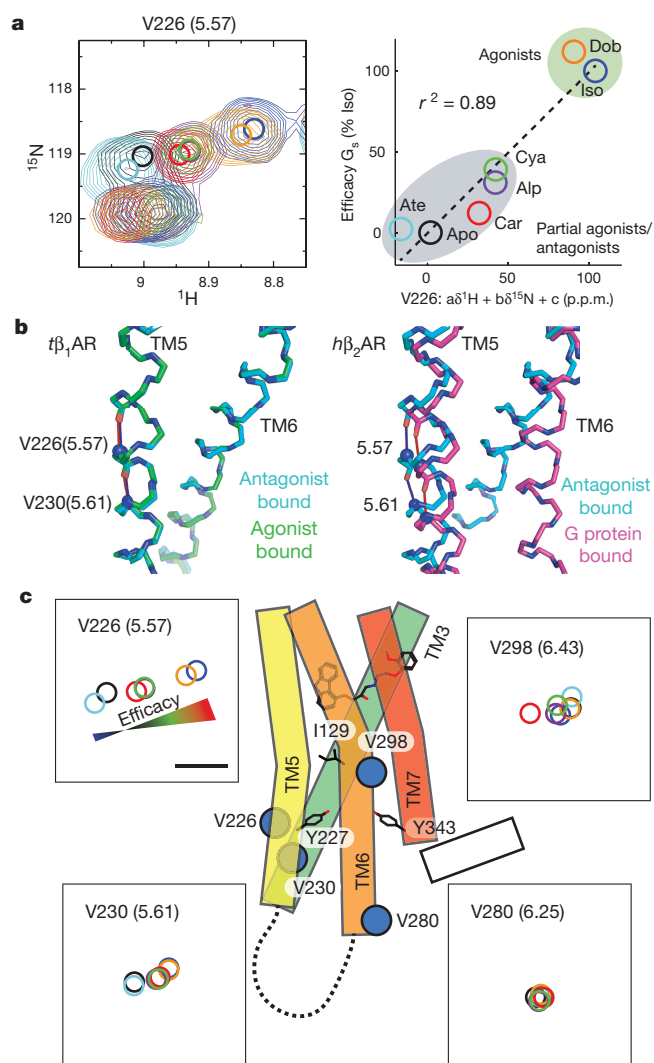
indicates that this extracellular part of the receptor undergoes micro-to millisecond motions in the apo form, which are quenched by ligand binding. This is consistent with results on  $\beta_2\text{AR}$ , which suggest that high-affinity ligands stabilize the conformation of ECL2 and ECL3<sup>5</sup>. The ligand-induced shifts of the V314 resonance correlate strongly ( $r^2 = 0.95$ ) with the reported ligand affinity<sup>17</sup> (Fig. 1d). Interestingly, the resonances of V314(6.59) in the apo form and in high-affinity ligand complexes are very close. This may indicate that the high-affinity ligand complexes mimic the average apo conformation. Finally, the  $^1\text{H}$ - $^{15}\text{N}$  chemical shifts of V125(3.36) at the bottom of the binding site and of V103(2.65) close to the ligand tail reveal additional trends (Fig. 1e): the chemical shifts of V125(3.36) correlate with the depth of ligand insertion towards the central part of TM3 ( $r^2 = 0.81$ ), whereas those of V103 correlate with the volume of the ligand tail ( $r^2 = 0.90$ ).

Compared to inactive  $\beta_2\text{AR}$ , complexes of activated  $\beta_2\text{AR}$  with either G protein<sup>2</sup> or the G protein-mimicking nanobody NB80<sup>9,18</sup> show large movements at the intracellular sides of TM5, TM6 and their intervening loop ICL3, which form the binding site for the G protein. These conformational changes are expected to be conserved throughout the GPCR family<sup>19</sup>. Four valine residues could be assigned in this region of TS- $\beta_1\text{AR}$ : V226(5.57), V230(5.61), V280(6.25), and V298(6.43) (Fig. 2). In contrast to the chemical shift changes in the vicinity of the ligand binding pocket, which depend strongly on the ligand chemistry, the shifts of the TM5 residues observed in this region report on ligand

efficacy. This effect is most prominent for residue V226(5.57), for which the  $^1\text{H}$ - $^{15}\text{N}$  resonances fall on one line from antagonists to agonists (Fig. 2a). The chemical shifts for the different ligands correlate very strongly ( $r^2 = 0.89$ ) with their reported<sup>17</sup> efficacies for Gs signalling (Extended Data Table 1). This highly linear effect suggests that the receptor filters the diverse input signals from the various ligands to a unified and precise structural response on TM5, which can be read out by the chemical shifts of V226(5.57). Interestingly, the V226(5.57) atenolol peak is situated at a position corresponding to lower efficacy than for the apo receptor. This gives direct structural evidence of atenolol's inverse agonist action, which reduces the activation relative to the basal level of the apo receptor.

Current high-resolution structures of  $\beta_1\text{AR}$  do not show significant changes between antagonist- and agonist-bound forms (Fig. 2b). The decrease of the V226(5.57)  $^1\text{H}$ - $^{15}\text{N}$  chemical shift by about 0.2 p.p.m. from the agonist isoprenaline to the antagonist atenolol indicates a lengthening of the V226(5.57)-H<sup>N</sup>...I222(5.53)-O hydrogen bond by about 0.05 Å (ref. 20). This small, but clearly NMR-detectable length variation is below the resolution limit of current GPCR structures (Supplementary Text 1), but may indicate the start of TM5 bending towards the active conformation as observed in the G protein-bound form of  $\beta_2\text{AR}$  (Fig. 2b). Remarkably, this response to agonists occurs even in the thermostabilized receptor TS- $\beta_1\text{AR}$ . Albeit reduced in absolute size compared to V226(5.57), V230(5.61) displays similar linear chemical shift changes as a function of ligand efficacy for the





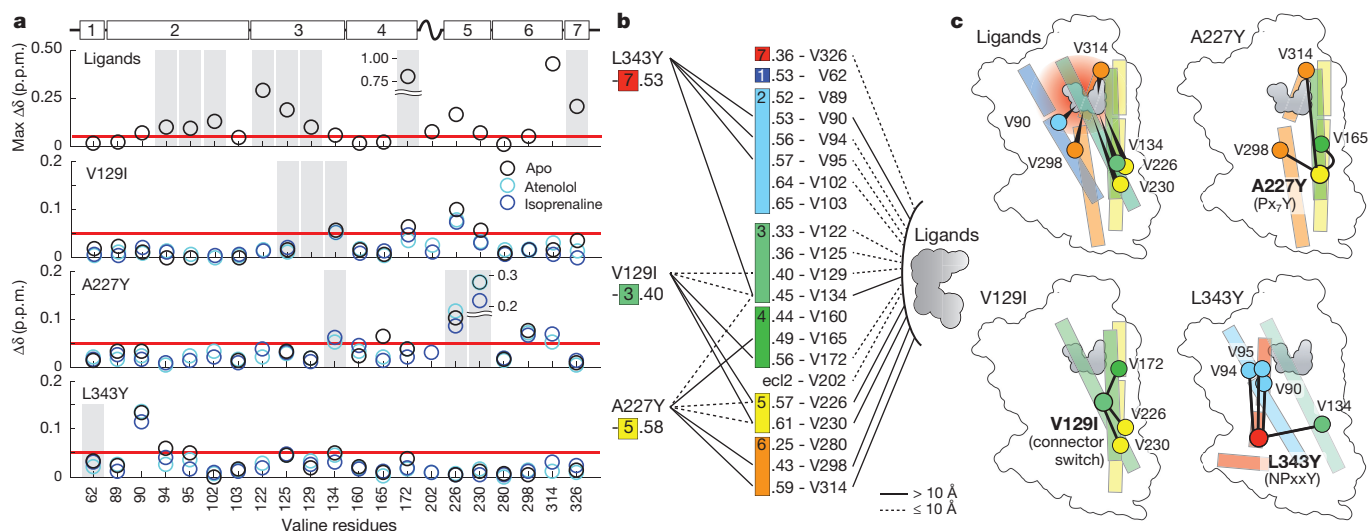
**Figure 2 | Correlation of ligand-induced chemical shift changes at the TS- $\beta_1$ AR intracellular side with  $G_s$  efficacy.** **a**, Left, response of the V226(5.57)  $^1\text{H}$ - $^{15}\text{N}$  resonance to various ligands (colour coding as in Fig. 1). The centres of resonances are indicated by circles. The  $^1\text{H}$ - $^{15}\text{N}$  resonances fall on one line from atenolol (antagonist) over apo to alprenolol (partial agonist), carvedilol (antagonist), cyanopindolol (partial agonist), dobutamine (full agonist) and isoprenaline (full agonist). **a**, Right, correlation of a best-fit linear combination of the V226(5.57) chemical shifts ( $-515 \delta^1\text{H} - 31.7 \delta^{15}\text{N} + 8.41 \cdot 10^3$ ) in different ligand complexes to their efficacy for the  $G_s$  signalling pathway<sup>17</sup>. **b**, Left, overlay of TM5 and TM6 backbones of thermostabilized  $\beta_1$ AR in antagonist- (blue, PDB code 4AMJ) and agonist-bound (green, PDB code 2Y03) form. The agonist does not induce detectable helix movements. **b**, Right, TM5 and TM6 backbone movements upon activation in human  $\beta_2$ AR. The overlay of inactive (blue, PDB code 2RH1) and G protein-bound  $\beta_2$ AR (magenta, PDB code 3SN6) structures shows the large bend of TM6 along with the smaller conformational change of TM5 upon activation. Hydrogen bonds 5.57- $\text{H}^{\text{N}} \cdots 5.53\text{-O}$  and 5.61- $\text{H}^{\text{N}} \cdots 5.57\text{-O}$  are indicated by dashes. According to the behaviour of the  $^1\text{H}$  chemical shifts of V226(5.57) and V230(5.61) in TS- $\beta_1$ AR, these hydrogen bonds expand in an efficacy-dependent manner during agonist binding. **c**, Response of the  $^1\text{H}$ - $^{15}\text{N}$  resonances for V226(5.57), V230(5.61), V298(6.43), and V280(6.25) to various ligands. For clarity, only the centres of resonances are depicted as circles with colour coding as in Fig. 1. The black bar represents a scale of 0.1 p.p.m. and 1 p.p.m. for the  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts, respectively. The schematic representation of the receptor indicates the locations of the respective valine residues at the cytoplasmic sides of TM5 and TM6 within the helical bundle of  $\beta_1$ AR.

$G_s$  pathway (Fig. 2c). As this residue is located one helical turn further towards the cytoplasm, the detected conformational change is not just local, but spans a certain length in TM5.

Compared to V226(5.57) and V230(5.61) in TM5, the chemical shift response to ligands is much less pronounced for V298(6.43) and in particular V280(6.25) at the intracellular side of TM6 (Fig. 2c). This suggests that agonist binding to the TS- $\beta_1$ AR does not induce the large conformational change in TM6 observed in the activated  $\beta_2$ AR-G protein<sup>2</sup> or  $\beta_2$ AR-NB80<sup>18</sup> complexes. However, G protein activation upon agonist binding has been reported for other less thermostabilized  $\beta_1$ AR constructs<sup>17</sup>, indicating that they can still be activated, albeit at low levels. Thus we reverted the mutations most likely to interfere with the activation mechanism in TS- $\beta_1$ AR, that is, I129(3.40)V in the connector switch<sup>15</sup>, Y227(5.58)A in TM5<sup>21,22</sup>, and Y343(7.53)L in the NPxxY motif of TM7<sup>14,22</sup> to the native residues. These reverse mutants were then tested for G protein activation and the NMR response in TM6. A summary of the results is given in Extended Data Table 3. The original TS- $\beta_1$ AR, the single mutants TS- $\beta_1$ AR(V129I), TS- $\beta_1$ AR(A227Y), and TS- $\beta_1$ AR(L343Y) as well as the double mutant TS- $\beta_1$ AR(V129I/A227Y) showed no detectable G protein activation upon isoprenaline binding (Extended Data Fig. 5). However, G protein activation was detectable for the least thermostable TS- $\beta_1$ AR(A227Y/L343Y) double mutant ( $T_m$  reduced by 11 °C relative to TS- $\beta_1$ AR), which recovers the conserved tyrosines in TM5 and TM7 that are known to stabilize the active state of rhodopsin<sup>22</sup>. None of the reverse mutants showed major changes in the NMR spectra of various ligand complexes compared to the original TS- $\beta_1$ AR (Extended Data Fig. 6). In particular, residues V298(6.43) and V280(6.25) at the intracellular side of TM6 did not show an increased response to agonists. This is in agreement with recent DEER (double electron-electron resonance) and  $^{19}\text{F}$ -NMR data showing that agonists alone do not fully stabilize the active state of TM6 in  $\beta_2$ AR<sup>9</sup>. Interestingly, the V129(3.40)I and the V129(3.40)I/A227(5.58)Y mutations shifted the  $^1\text{H}$ - $^{15}\text{N}$  resonances of V226(5.57) towards a more active (that is, bent) state of TM5 in both the atenolol- and isoprenaline-bound forms (Extended Data Fig. 6b), thereby given direct experimental evidence for an allosteric activation pathway spanning about 13 Å from I129(3.40) on TM3 to V226(5.57) on TM5.

With the exception of rhodopsin<sup>23</sup>, the stabilization of fully active GPCR conformations seems to require binding of an agonist and an intracellular partner<sup>8,9</sup>. Indeed, when both the agonist isoprenaline and the G protein-mimicking nanobody NB80<sup>18</sup> were added to TS- $\beta_1$ AR(A227Y/L343Y), very large chemical shift responses for many valine residues in TM3–TM6 were observed, whereas no change was observed for several valines in TM1, 2, and 7 (Extended Data Fig. 7). This very strong response extends even to the extracellular residue V314(6.59), providing evidence of a long-distance connection from the G protein binding site to the ligand entry site. The strong chemical shift changes are reverted when the partial agonist/antagonist cyanopindolol is added to the isoprenaline-TS- $\beta_1$ AR(A227Y/L343Y)–NB80 complex. The spectrum then becomes identical to that of the ‘pure’ cyanopindolol-TS- $\beta_1$ AR(A227Y/L343Y) complex (Extended Data Fig. 7a), indicating that cyanopindolol replaces isoprenaline and causes the release of NB80. In agreement with the G protein activation data, the isoprenaline-bound original TS- $\beta_1$ AR and the mutants TS- $\beta_1$ AR(A227Y) and TS- $\beta_1$ AR(L343Y) did not show binding of NB80 in the NMR spectra. Moreover, supplementing NB80 to the ultra-stable TS- $\beta_1$ AR did not change its affinity for isoprenaline (Extended Data Fig. 7b), whereas it caused a hundred-fold affinity increase in the case of TS- $\beta_1$ AR(A227Y/L343Y) and the truncated native turkey  $\beta_1$ AR receptor (t $\beta_1$ trunc)<sup>17</sup>. This increase is identical to data for  $\beta_2$ AR<sup>18</sup> and shows the energetic coupling between the NB80 and agonist binding also for  $\beta_1$ AR.

In combination, these data prove that agonist binding, even in the absence of a G protein mimic, induces initial changes in the conformational equilibrium of TM5 towards the conformation observed in the G protein complex of  $\beta_2$ AR. Remarkably, these rearrangements occur in all thermostabilized forms of  $\beta_1$ AR. However, a full shift of the equilibrium towards such an active conformation, including allosteric



**Figure 3 | Experimental detection of allosteric signalling pathways using the NMR response to ligand binding and point mutations at different backbone sites.** **a**, Combined  $^1\text{H}$ ,  $^{15}\text{N}$  chemical shift deviations [ $\Delta\delta = (\Delta\delta_{\text{H}}^2/2 + \Delta\delta_{\text{N}}^2/50)^{1/2}$ ] of valine resonances observed upon ligand binding or induced by the indicated point mutations. For ligand binding, the three pairwise deviations  $\Delta\delta$  were calculated between the apo, atenolol-, and isoprenaline-bound forms of TS- $\beta_1$ AR. The maximum of these deviations is shown. For the reverse mutants, deviations  $\Delta\delta$  are shown relative to TS- $\beta_1$ AR for their apo (black), atenolol- (cyan), and isoprenaline-bound (blue) bound forms. Valines within 10 Å from the C $\alpha$  atom of the mutated amino acid are shown on a grey background. Distances were calculated using the coordinates of the thermostabilized  $\beta_1$ AR (PDB code 4BVN). A red line marks a cut-off value  $\Delta\delta$  of 0.05 p.p.m. for significant chemical shift deviations. **b**, Topology of the signalling

network determined from point mutations (left) and ligand binding (right). Signal paths were identified by chemical shift deviations  $\Delta\delta$  larger than 0.05 p.p.m. induced by these two perturbations (**a**). Signal paths to valines within 10 Å from the ligand or point mutation (that is, localized conformational changes) are indicated as dashed lines, and those beyond 10 Å (long-range conformational changes) as solid lines. The ligand signals broadly towards all helices but TM1. In contrast, the network determined by the point mutations is more localized and connects TM3 to TM4/5, TM5 to TM3/4/6, and TM7 to TM2/3. The latter network seems to be divided into two subnetworks involving TM3/4/5/6 and TM2/7. **c**, Long-range allosteric signal paths identified from ligand binding or point mutations (**a** and **b**) indicated on schematic  $\beta_1$ AR representations showing the involved TMs. Helices are colour-coded according to **b**.

changes at the extracellular side, occurs only when G protein or its mimetic NB80 is bound. This process requires the presence of both Y227(5.58) and Y343(7.53), which significantly reduce the thermal stability. Different active conformations may be reached for non-G protein effectors such as  $\beta$ -arrestin.

The possibility to detect NMR signals at many receptor sites in response to ligand binding and point mutations provides an experimental method to trace allosteric signalling paths. Figure 3 shows examples of these pathways, derived from the response to the ligands atenolol and isoprenaline and the single point mutations V129(3.40)I, A227(5.58)Y, and L343(7.53)Y. Choosing a cutoff of 0.05 p.p.m. for the resulting combined  $^1\text{H}$ ,  $^{15}\text{N}$  chemical shift change (Fig. 3a, red line), long-range (> 10 Å) connections become evident throughout the receptor (Fig. 3b, c). Whereas detected ligand signals radiate broadly to almost all helices, the point mutants give evidence of smaller interaction networks connecting TM3 to TM4/5, TM5 to TM3/4/6, as well as TM7 to TM2/3 (Fig. 3b, c). Interestingly, the TM2/TM7 network seems to be only weakly connected to the TM3–TM6 network. Together, these data provide experimental evidence at high resolution of an extensive signal transduction network that connects the ligand binding site to the intracellular sides of TM5, TM6, and TM7. Such a network of loosely coupled allosteric connections has been postulated previously for  $\beta_2$ AR on the basis of molecular dynamics simulations<sup>24</sup>.

In summary, we have shown that highly resolved solution NMR backbone spectra can be obtained for a eukaryotic GPCR. The NMR observations delineate the allosteric signalling pathways and comprehensively connect many previous experimental and theoretical observations, which may ultimately allow to understand the dynamic mechanisms of GPCRs at the atomic level.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 October 2014; accepted 4 December 2015.

Published online 3 February 2016.

- Rasmussen, S. G. F. *et al.* Crystal structure of the human  $\beta_2$  adrenergic G-protein-coupled receptor. *Nature* **450**, 383–387 (2007).
- Rasmussen, S. G. F. *et al.* Crystal structure of the  $\beta_2$  adrenergic receptor-Gs protein complex. *Nature* **477**, 549–555 (2011).
- Ghosh, E., Kumari, P., Jaiman, D. & Shukla, A. K. Methodological advances: the unsung heroes of the GPCR structural revolution. *Nature Rev. Mol. Cell Biol.* **16**, 69–81 (2015).
- Grzesiek, S. & Sass, H.-J. From biomolecular structure to functional understanding: new NMR developments narrow the gap. *Curr. Opin. Struct. Biol.* **19**, 585–595 (2009).
- Bokoch, M. P. *et al.* Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor. *Nature* **463**, 108–112 (2010).
- Kofuku, Y. *et al.* Efficacy of the  $\beta_2$ -adrenergic receptor is determined by conformational equilibrium in the transmembrane region. *Nat. Commun.* **3**, 1045–1049 (2012).
- Liu, J. J., Horst, R., Katritch, V., Stevens, R. C. & Wüthrich, K. Biased signaling pathways in  $\beta_2$ -adrenergic receptor characterized by  $^{19}\text{F}$ -NMR. *Science* **335**, 1106–1110 (2012).
- Nygaard, R. *et al.* The dynamic process of  $\beta_2$ -adrenergic receptor activation. *Cell* **152**, 532–542 (2013).
- Manglik, A. *et al.* Structural insights into the dynamic process of  $\beta_2$ -adrenergic receptor signaling. *Cell* **161**, 1101–1111 (2015).
- Warne, T. *et al.* Structure of a  $\beta_1$ -adrenergic G-protein-coupled receptor. *Nature* **454**, 486–491 (2008).
- Miller, J. L. & Tate, C. G. Engineering an ultra-thermostable  $\beta_1$ -adrenoceptor. *J. Mol. Biol.* **413**, 628–638 (2011).
- Miller-Gallacher, J. L. *et al.* The 2.1 Å resolution structure of cyanopindolol-bound  $\beta_1$ -adrenoceptor identifies an intramembrane  $\text{Na}^+$  ion that stabilises the ligand-free receptor. *PLoS One* **9**, e92727 (2014).
- Ballesteros, J. A. & Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci.* **25**, 366–428 (1995).
- Deupi, X. & Standfuss, J. Structural insights into agonist-induced activation of G-protein-coupled receptors. *Curr. Opin. Struct. Biol.* **21**, 541–551 (2011).
- Dror, R. O. *et al.* Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl Acad. Sci. USA* **108**, 13118–13123 (2011).

16. González, A., Perez-Acle, T., Pardo, L. & Deupi, X. Molecular basis of ligand dissociation in  $\beta$ -adrenergic receptors. *PLoS One* **6**, e23815 (2011).
17. Baker, J. G., Proudman, R. G. W. & Tate, C. G. The pharmacological effects of the thermostabilising (m23) mutations and intra and extracellular ( $\beta$ 36) deletions essential for crystallisation of the turkey  $\beta$ -adrenoceptor. *Naunyn Schmiedebergs Arch. Pharmacol.* **384**, 71–91 (2011).
18. Rasmussen, S. G. F. *et al.* Structure of a nanobody-stabilized active state of the  $\beta_2$  adrenoceptor. *Nature* **469**, 175–180 (2011).
19. Manglik, A. & Kobilka, B. The role of protein dynamics in GPCR function: insights from the  $\beta_2$ AR and rhodopsin. *Curr. Opin. Cell Biol.* **27**, 136–143 (2014).
20. Grzesiek, S., Cordier, F., Jaravine, V. & Barfield, M. Insights into biomolecular hydrogen bonds from hydrogen bond scalar couplings. *Prog Nucl Mag Res Sp* **45**, 275–300 (2004).
21. Tate, C. G. & Schertler, G. F. Engineering G protein-coupled receptors to facilitate their structure determination. *Curr. Opin. Struct. Biol.* **19**, 386–395 (2009).
22. Goncalves, J. A. *et al.* Highly conserved tyrosine stabilizes the active state of rhodopsin. *Proc. Natl Acad. Sci. USA* **107**, 19861–19866 (2010).
23. Park, J. H., Scheerer, P., Hofmann, K. P., Choe, H.-W. & Ernst, O. P. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* **454**, 183–187 (2008).
24. Dror, R. O. *et al.* Activation mechanism of the  $\beta_2$ -adrenergic receptor. *Proc. Natl Acad. Sci. USA* **108**, 18684–18689 (2011).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We acknowledge T. Sharpe for expert help with biophysical assays. We thank J. Steyaert for providing the NB80 plasmid and M. Rogowski for preparing wild-type  $\beta_1$ AR baculovirus. F.B. was supported by Marie Curie and EMBO postdoctoral fellowships. This work was supported by Swiss National Science Foundation grants 31-132857 (S.G.), Sinergia 141898 (S.G., D.B.V., G.F.X.S.), 31-135754 (D.B.V.), 31-153145 (G.F.X.S.), 31-146520 (X.D.), European Union FP7 grant 242135 (S.G.), and COST Action CM1207 (GLISTEN) (X.D. and G.F.X.S.).

**Author Contributions** S.G., G.F.X.S., D.B.V., X.D. and S.I. initiated and managed the project. S.I., X.D., F.B., D.B.V., G.F.X.S. and S.G. designed ligand response and selective labelling experiments. C.O. designed the initial TS- $\beta_1$ AR construct and established purification. S.I. designed and prepared all selectively labelled receptor mutants, performed ligand exchange, NB80 binding, and all NMR experiments. S.I. and S.G. designed NMR experiments, analysed and interpreted all data. F.M.H. and D.B.V. purified trimeric  $G_i$  protein, designed and performed radioligand affinity and  $G_i$  protein activation assays. C.-J.T. prepared NB80 nanobody. S.G., S.I., X.D., F.M.H., D.B.V. and G.F.X.S. wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.G. ([stephan.grzesiek@unibas.ch](mailto:stephan.grzesiek@unibas.ch)), D.B.V. ([dmitry.veprintsev@psi.ch](mailto:dmitry.veprintsev@psi.ch)) or G.F.X.S. ([gebhard.schertler@psi.ch](mailto:gebhard.schertler@psi.ch)).



## METHODS

**$\beta_1$ AR constructs.** The TS- $\beta_1$ AR mutant was derived from the turkey  $\beta_1$ AR44-m23 mutant used in crystallographic studies<sup>25</sup> by adding three additional thermostabilizing mutations (I129V, D322K, and Y343L) and a neutral mutation (D200E) from the ultra-stable  $\beta_1$ AR-JM3 mutant<sup>11</sup>. As compared to the wild type, TS- $\beta_1$ AR contains truncations at the amino and carboxy termini and intracellular loop (ICL3), a total of nine thermostabilizing point mutations, three further point mutations as well as a C-terminal hexahistidine tag (Extended Data Fig. 1). The final TS- $\beta_1$ AR sequence is MGAELLSQQWEAGMSLLMALVLLVAGNVLVIAAIGSTQRLQTLTNLFITSLACADLVVGLVVPFGATLVVRGTWLVGWSFLCELWTSLDVLCVTASVETLCVIAIDRYLAITSPFRYQSLMTRARAKVIICTVWALSALVSFLPIMMHWWRDEDQALKCYQDPGCEFTVNRAYAIASSIIFYIPLLMIFVALRVYREAKEQIRKIDRASKRKTSRVMLMRHKKALGTGLIMGVFTLCWLPFLVNVNVTNFRDLVPKWLFFAVFNWLGYSANMNPILCRSPDFRKAFLKLLAFPRKADRRLLHHHHH.

Additional valine-to-alanine or isoleucine point mutations were introduced into TS- $\beta_1$ AR for NMR assignment purposes. All constructs were made using the QuikChange site-directed mutagenesis method (Agilent). Baculovirus for insect cell expression was generated using the Bac-to-Bac system (Invitrogen).

**$\beta_1$ AR expression and purification.** All  $\beta_1$ AR constructs were expressed in baculovirus-infected insect cells as described<sup>26</sup>. Selective labelling by [<sup>15</sup>N]valine was achieved by growing cells on unlabelled serum-free insect cell medium (InsectXpress, Lonza) and then exchanging into custom-made serum-free medium (SF4, BioConcept) devoid of valine and yeast extract, to which 100 mg l<sup>-1</sup> [<sup>15</sup>N]valine were supplemented. Virus was added immediately after the medium exchange. The culture was harvested at 48 or 72 h post infection.

After cell lysis, the membrane fraction was separated from the lysate via ultracentrifugation and subsequently solubilized with 2% *n*-decyl- $\beta$ -D-maltopyranoside (DM, Anatrace). The solubilized membrane fraction was then purified by nickel ion affinity chromatography followed by alprenolol ligand affinity chromatography. The active receptor was eluted with buffer (20 mM Tris, 350 mM NaCl, 0.1% DM, pH 7.5) containing either atenolol (1 mM) or alprenolol (0.1 mM). Final yields of detergent-solubilized receptor were 1.5 mg l<sup>-1</sup> of cell culture. The molecular weight of the receptor-detergent complex was estimated as ~100 kDa by static light scattering.

**Thermal shift assays of mutant receptors.** Detergent-solubilized, purified apo TS- $\beta_1$ AR and reverse-mutation receptors for thermal stability assays were obtained from their atenolol-bound form by washing with buffer devoid of ligand on a HiTrap SP HP (GE Healthcare) column. Their thermal stability was determined by the microscale fluorescent stability assay for binding of the thiol-specific fluorochrome N-[4-(7-diethylamino-4-methyl-3-coumarinyl)phenyl]maleimide (CPM)<sup>27</sup> in a Rotor-Gene Q (QIAGEN) real-time PCR cycler using 1  $\mu$ g of receptor in 20 mM Tris, 350 mM NaCl, 0.1% DM, pH 7.5 and a heating rate of 2 K min<sup>-1</sup>.

**NMR experiments.** NMR samples were prepared in Shigemi tubes as 250  $\mu$ l volumes of typically 100  $\mu$ M receptor, 1 mM ligand (except for apo form), 20 mM TRIS, 100 mM NaCl, 0.1% DM, 5% D<sub>2</sub>O, pH 7.5. For isoprenaline or dobutamine, 2 mM of sodium L-ascorbate were supplemented as anti-oxidant. All solution NMR measurements were carried out at 304 K on a 800 MHz or a 900 MHz Bruker Avance III spectrometer equipped with a cryogenic probe. 2D <sup>1</sup>H, <sup>15</sup>N TROSY spectra were recorded with total acquisition periods of 16 ms (<sup>15</sup>N) and 43 ms (<sup>1</sup>H) with typical total experimental times of 24–48 h. As compared to a standard TROSY pulse sequence, the <sup>1</sup>H-<sup>15</sup>N INEPT delays were set to 3 ms to reduce magnetization losses from relaxation.

**Assignment procedure.** To obtain sequence-specific assignment information, we initially attempted to detect HNCO and HNCA correlations on samples additionally labelled with <sup>13</sup>C at specific backbone sites<sup>28</sup>. However, due to low sensitivity, only very few correlations were observable. Therefore, assignments were obtained from a combination of information from TROSY spectra recorded on 18 TS- $\beta_1$ AR valine point mutants with different ligands, four HN(CO) correlations and five distinct structure-based chemical shift predictions (Extended Data Table 2).

**Ligand exchange experiments.** Receptor complexes with different ligands were generated by sequential exchange according to increasing ligand affinity, that is, in the sequence atenolol-isoprenaline-dobutamine-alprenolol or alprenolol-carvedilol-cyanopindolol. For exchange, the sample was washed three times with buffer devoid of ligand at tenfold dilution in Amicon Ultra 50 kDa cutoff concentrators. Subsequently, the sample was washed again twice with buffer containing 100  $\mu$ M new ligand, separated by a period of 1 h incubation. Final concentrations of the ligands were adjusted to 1 mM. Apo receptor was generated from the atenolol

complex by six washing steps of tenfold dilution in ligand-free buffer using a 1 h incubation period for the last three steps.

**NMR NB80 binding experiment.** Binding of NB80 to  $\beta_1$ AR mutants was assessed using TROSY and 1D proton NMR spectra. These spectra were recorded on the  $\beta_1$ AR mutants (TS- $\beta_1$ AR: 132  $\mu$ M, TS- $\beta_1$ AR(A227Y): 120  $\mu$ M, TS- $\beta_1$ AR(L343Y): 110  $\mu$ M, and TS- $\beta_1$ AR(A227Y/L343Y): 120  $\mu$ M) in the presence of saturating amounts (1 mM) of the agonist isoprenaline before and immediately after addition of an equimolar (relative to the receptor) amount of NB80. For TS- $\beta_1$ AR(A227Y/L343Y) additional spectra were recorded after a further addition of the partial agonist cyanopindolol (1 mM) to the already present isoprenaline and NB80.

**Scintillation proximity assay with <sup>3</sup>H-dihydroalprenolol.** For pharmacological binding assays membranes were prepared from SF9 insect cells as described previously<sup>26</sup>. The total protein content of the membranes was estimated by A280 measurements using an average extinction coefficient of 1.0 per mg ml<sup>-1</sup>. All assays were carried out in 96-well plates at 200  $\mu$ g ml<sup>-1</sup> total protein in membranes and 2 mg ml<sup>-1</sup> WGA-Ysi beads (Perkin-Elmer) in a 100  $\mu$ l total volume per well. Samples were equilibrated at room temperature for at least 16 h. *K<sub>D</sub>* values for the radioactive ligand [<sup>3</sup>H]dihydroalprenolol (<sup>3</sup>H-DHA) were determined by titrating <sup>3</sup>H-DHA from 0.032 to 100 nM. Non-specific binding was determined in presence of 1  $\mu$ M S-propanolol to block the ligand binding site. Competition assays were performed in the presence of 20 nM <sup>3</sup>H-DHA (hot ligand) and increasing concentrations of the competitor (cold ligand). Dilutions of alprenolol, atenolol, cyanopindolol, dobutamine and isoprenaline were made with phosphate buffered saline (PBS, Biochrom, Germany). Due to the limited solubility of carvedilol in water, stock dilutions of carvedilol were prepared in DMSO. The final concentration of DMSO in the samples was 5%. To test the effect of NB80 binding on the affinities (IC<sub>50</sub>) of isoprenaline for various receptor mutants, the competition assays were also carried out in the presence of saturating concentrations of NB80 (10  $\mu$ M). IC<sub>50</sub> values were determined by fitting the measured radioactive counts per minute CPM(X) at a specific concentration X of the competitor

to the equation 
$$\text{CPM}(X) = \frac{\text{CPM}_{\text{max}} - \text{CPM}_{\text{min}}}{1 + \frac{X}{\text{IC}_{50}}} + \text{CPM}_{\text{min}}$$
, where CPM<sub>max</sub> and

CPM<sub>min</sub> are maximal and minimal counts of the assay, respectively. The fits were carried in MATLAB (MathWorks, <http://www.mathworks.com>) with Monte-Carlo estimation of errors. *K<sub>i</sub>* values were calculated from the obtained IC<sub>50</sub>

values according to the formula 
$$K_i = \frac{\text{IC}_{50}}{1 + \frac{A}{K_D}}$$
 where A is the concentration of the

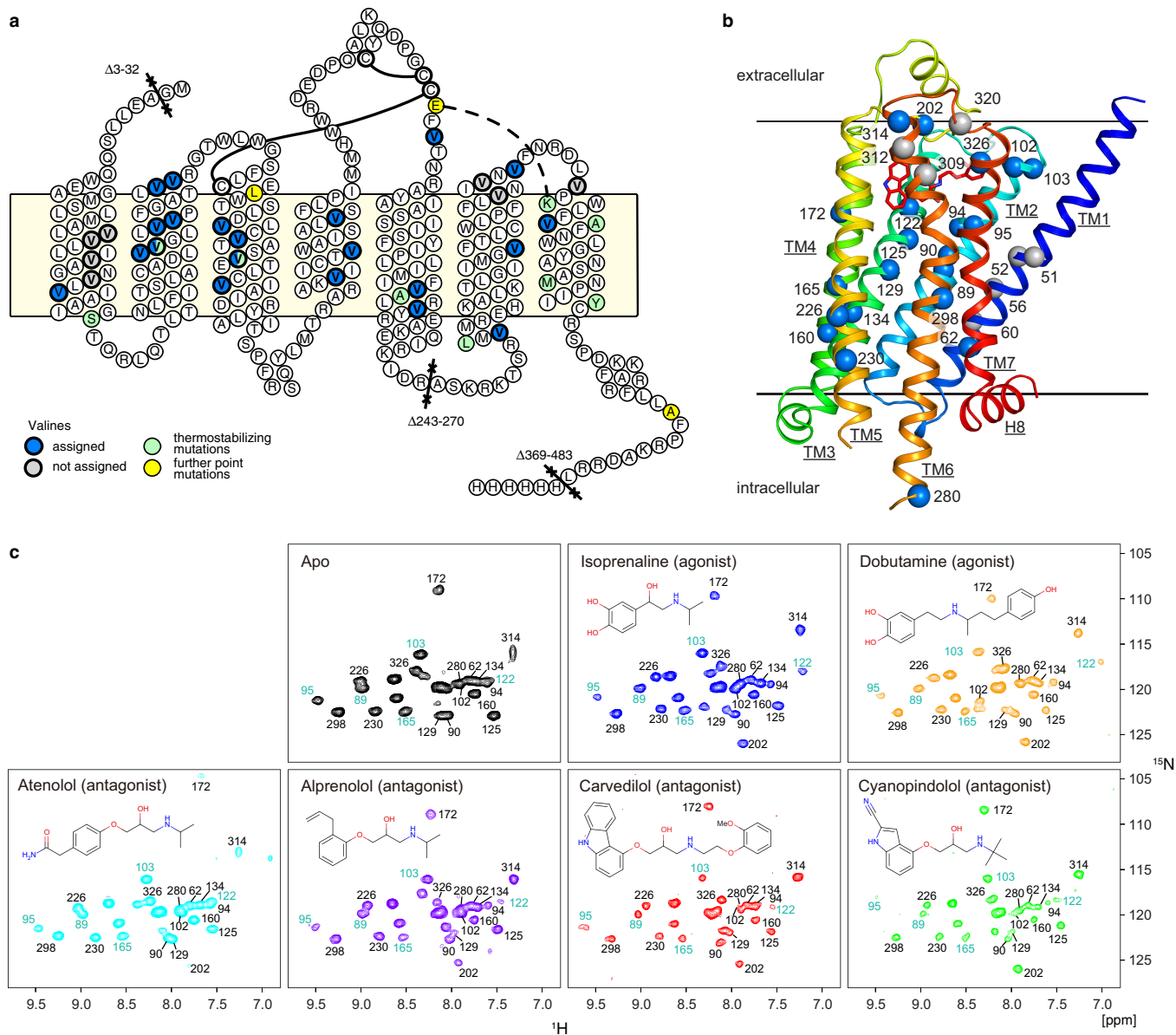
radioactive ligand and *K<sub>D</sub>* is its affinity for the receptor determined in the direct binding experiment.

**G protein activation assay.** G protein activation was measured on purified  $\beta_1$ AR mutants reconstituted with MSP1E3D1<sup>29</sup> into POPC/POPG nanodiscs. MSP1E3D1 was expressed and purified as described<sup>29</sup> and cleaved with TEV protease. 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC, Avanti Polar Lipids) and 1-palmitoyl-2-oleoyl-*sn*-glycerol-3-phospho-(1'-*rac*-glycerol) sodium salt (POPG, Avanti Polar Lipids) were solubilized at a ratio of 1:1.5 (w/w) POPG/POPC in ND buffer (20 mM HEPES pH 8, 100 mM NaCl, 1 mM EDTA) with 50 mM sodium cholate (Sigma-Aldrich) at 4°C. 133.3  $\mu$ M MSP1E3D1 was incubated with 8 mM solubilized POPC/POPG and 10  $\mu$ M purified  $\beta_1$ AR in ND buffer with a final concentration of 24 mM sodium cholate for 1 h at 4°C. Nanodiscs containing the receptor were separated from empty nanodiscs using a cobalt-chelating resin. The heterotrimeric G protein was prepared by incubating 10  $\mu$ M recombinant G $\alpha_{i1}$  and 10  $\mu$ M native G $\beta\gamma_i$  in activation buffer (25 mM HEPES pH 7.5, 150 mM NaCl, 2 mM MgCl<sub>2</sub>, 1 mM DTT) for 30 min at 4°C.

G protein activation was detected by the change in tryptophan fluorescence caused by the exchange of GDP for GTP $\gamma$ S, associated conformational changes in the G $\alpha$  subunit and its dissociation from the G $\beta\gamma$  subunit of the heterotrimeric G protein<sup>30</sup>. All measurements were carried out on a Varian Cary Eclipse fluorescence spectrophotometer ( $\lambda_{\text{ex}}$  = 295 nm,  $\lambda_{\text{em}}$  = 340 nm, 1.5 nm excitation slit, 20 nm emission slit, 2 s averaging time, 15 s cycle time) using final sample volumes of 1 ml in 10  $\times$  4 mm cuvettes (Hellma, CH) and magnetic stirrers at 20°C. Prior to activation, the fluorescence intensity baseline was recorded with 100 nM heterotrimeric G protein for approximately 500 s. The activation was started by adding 6 nM  $\beta_1$ AR and 10  $\mu$ M GTP $\gamma$ S, and the fluorescence intensity was monitored for a further 1 h. For experiments in the presence of an agonist, the concentrated receptor stock solution (1.5  $\mu$ M) was pre-incubated for 30 min at 4°C with 40  $\mu$ M isoprenaline, and the buffer during the measurements contained 2  $\mu$ M of isoprenaline to maintain the saturation conditions for the receptor.

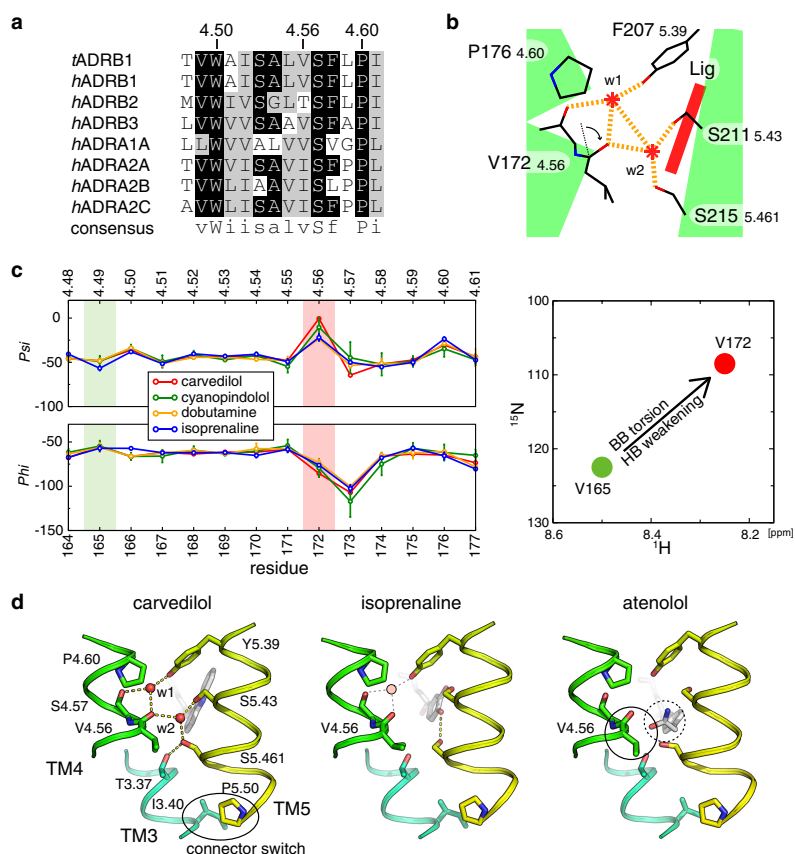


25. Warne, T. *et al.* The structural basis for agonist and partial agonist action on a  $\beta_1$ -adrenergic receptor. *Nature* **469**, 241–244 (2011).
26. Brueckner, F. *et al.* Structure of  $\beta$ -adrenergic receptors. *Methods Enzymol.* **520**, 117–151 (2013).
27. Alexandrov, A. I., Mileni, M., Chien, E. Y. T., Hanson, M. A. & Stevens, R. C. Microscale fluorescent thermal stability assay for membrane proteins. *Structure* **16**, 351–359 (2008).
28. Vajpai, N. *et al.* Backbone NMR resonance assignment of the Abelson kinase domain in complex with imatinib. *Biomol. NMR Assign.* **2**, 41–42 (2008).
29. Bayburt, T. H., Grinkova, Y. V. & Sligar, S. G. Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins. *Nano Lett.* **2**, 853–856 (2002).
30. Ernst, O. P., Bieri, C., Vogel, H. & Hofmann, K. P. Intrinsic biophysical monitors of transducin activation: fluorescence, UV-visible spectroscopy, light scattering, and evanescent field techniques. *Methods Enzymol.* **315**, 471–489 (2000).
31. Han, B., Liu, Y., Ginzinger, S. W. & Wishart, D. S. SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR* **50**, 43–57 (2011).



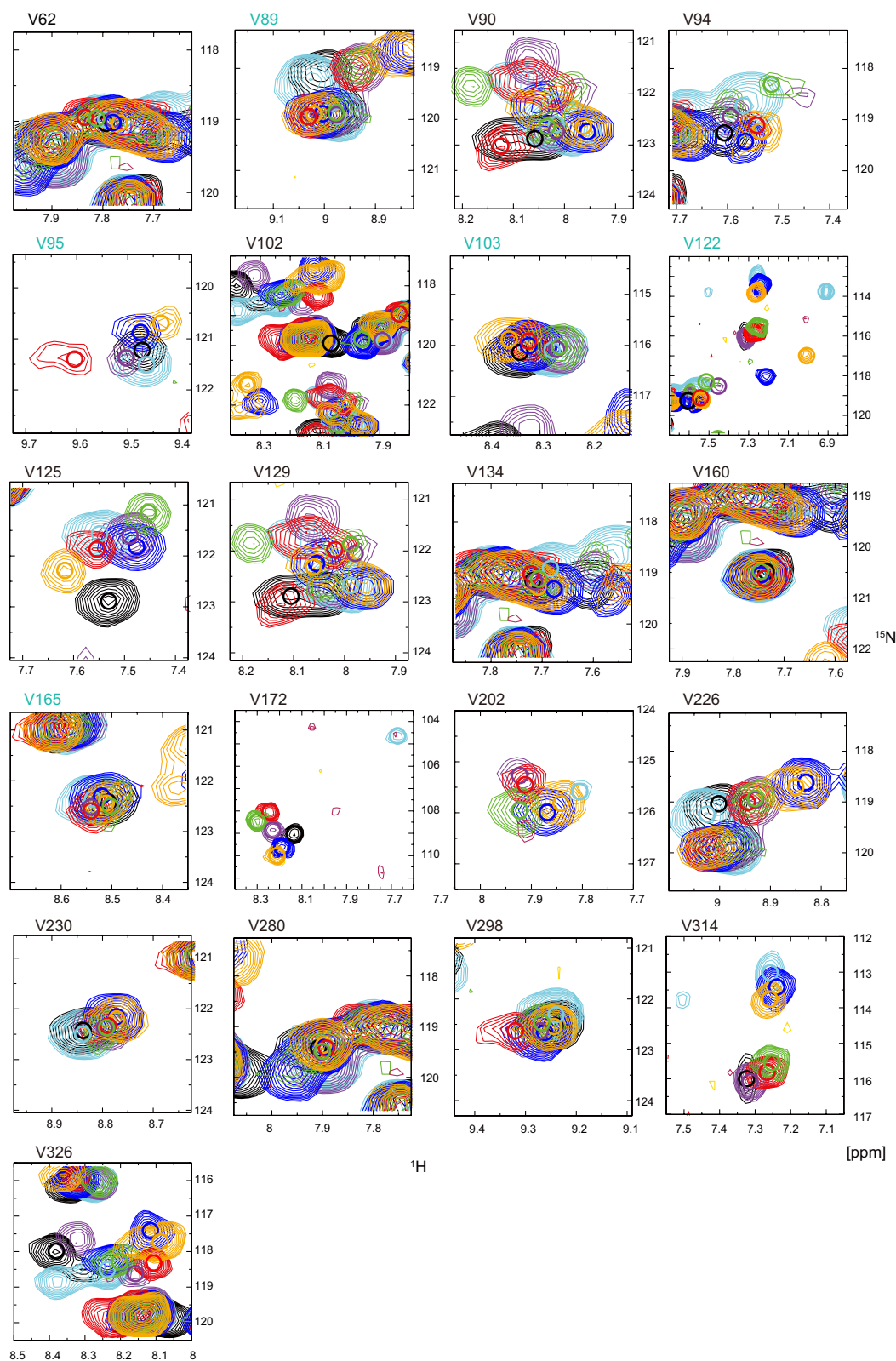
**Extended Data Figure 1 | Position of valine residues in the thermostabilized mutant TS- $\beta_1$ AR.** **a**, Schematic representation of secondary structure and amino acid sequence of TS- $\beta_1$ AR. As compared to the wild type, the TS- $\beta_1$ AR has truncations at the N and C termini and the intracellular loop (ICL3) as well as nine thermostabilizing point mutations (colour coded green) and three additional point mutations (colour coded yellow). Valines labelled with  $^{15}\text{N}$  are indicated by bold circles for assigned (blue) and unassigned (grey) residues. **b**, Structure of  $\beta_1$ AR in

complex with carvedilol (PDB code 4AMJ). The protein backbone and carvedilol are shown in ribbon and red stick representation, respectively. The individual valines are depicted as spheres (blue, assigned; grey, not assigned) labelled by residue number. c, Full  $^1\text{H}$ - $^{15}\text{N}$  TROSY spectra of apo TS- $\beta_1$ AR and all investigated ligand complexes. The ligand chemical structures are shown as inserts. Resonances are marked with assignment information (black, firm; cyan, tentative).



**Extended Data Figure 2 | Effect of ligand head group substitution on hydrogen bond network involving V172(4.56) in  $\beta_1$ AR.** **a**, Sequence alignment of turkey  $\beta_1$ AR and human adrenoreceptors around position 4.56 in TM4 showing the conserved proline at position 4.60. **b**, Schematic representation of the hydrogen bond network between TM4 and TM5 involving V172(4.56) in  $\beta_1$ AR. The hydrogen bond network (orange dotted lines) originates at the carbonyl of V172(4.56) and connects to TM5 via two water molecules (w1 and w2, red asterisks). **c**, Left, phi and psi dihedral angles (averages and standard deviations in case of multiple chains) in TM4 between residues 4.48 and 4.61 for different complexes of  $\beta_1$ AR (ligand, PDB code [chain]: carvedilol, 4AMJ [A,B]; cyanopindolol, 2VT4 [B,D] and 4BVN [A]; dobutamine, 2Y00 [A,B] and 2Y01 [A,B]; isoprenaline, 2Y03 [A,B]). V165(4.49) and V172(4.56), for which  $^1\text{H}$ - $^{15}\text{N}$  resonances could be observed, are shown on a green and red background, respectively. **c**, Right, average of the  $^1\text{H}$ - $^{15}\text{N}$  resonance positions for V165(4.49, green) and V172(4.56, red). The phi and psi values of V172(4.56) are distorted from the normal helical angles due to

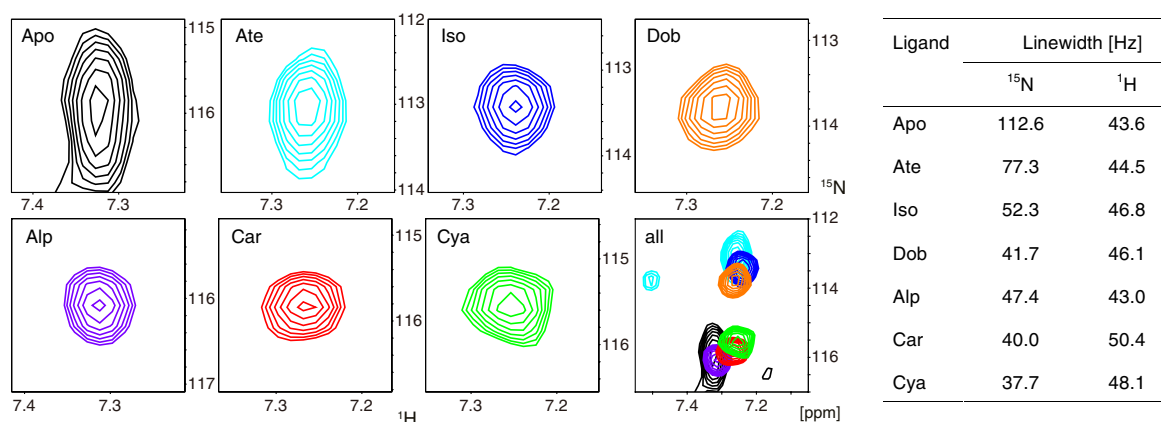
the loss of the intra-helical hydrogen bond to P176(4.60). Together with the loss of the canonical hydrogen bond, this strongly shifts both the  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts of V172(4.56) towards smaller p.p.m. values relative to V165(4.49), which has normal, helical phi and psi angles. **d**, Partial views of the crystal structures of the carvedilol complex (PDB code 4AMJ), the isoprenaline complex (PDB code 2Y03) and a docking model of the atenolol complex based on the cyanopindolol complex crystal structure (PDB code 4BVN). In the carvedilol complex, one water molecule (w2) forms a hydrogen bond network between the carbonyl oxygen of V172(4.56) and the side chains of S211(5.43) and S215(5.461). In the isoprenaline complex, the side chains of S211(5.43) and S215(5.461) are rotated and form hydrogen bonds to the catechol hydroxyl groups of isoprenaline. In this structure, no water molecule is observed at the equivalent position of w2. In the atenolol complex, the *para*-substituted acetamide of the ligand head ring (dashed circle) inserts between S211(5.43) and S215(5.461) and disturbs the interface between TM3 and TM5 near V172(4.56) (solid circle).



**Extended Data Figure 3 | Response to various ligands for all assigned valine  $^1\text{H}$ - $^{15}\text{N}$  resonances in TS- $\beta_1\text{AR}$ .** Colour coding as in Extended Data Fig. 1: cyan (atenolol), blue (isoprenaline), orange (dobutamine),

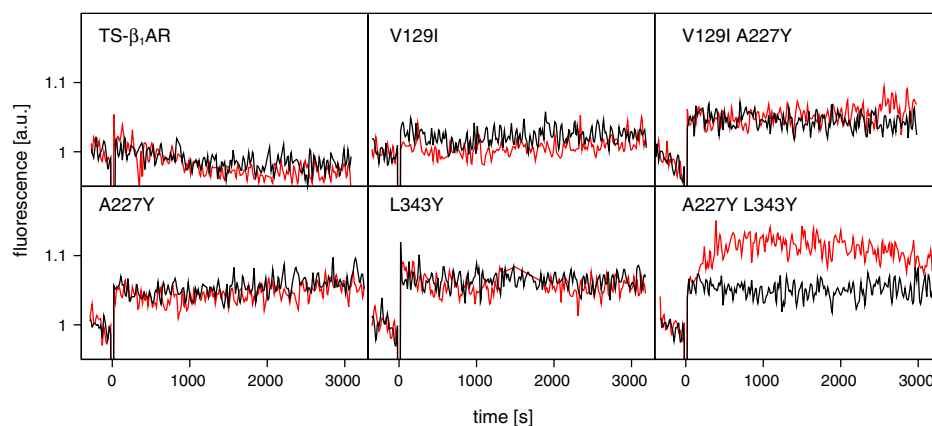
purple (alprenolol), red (carvedilol), green (cyanopindolol), black (apo receptor). For clarity the centres of resonances are marked by circles. Firmly (tentatively) assigned residues are marked in black (cyan).





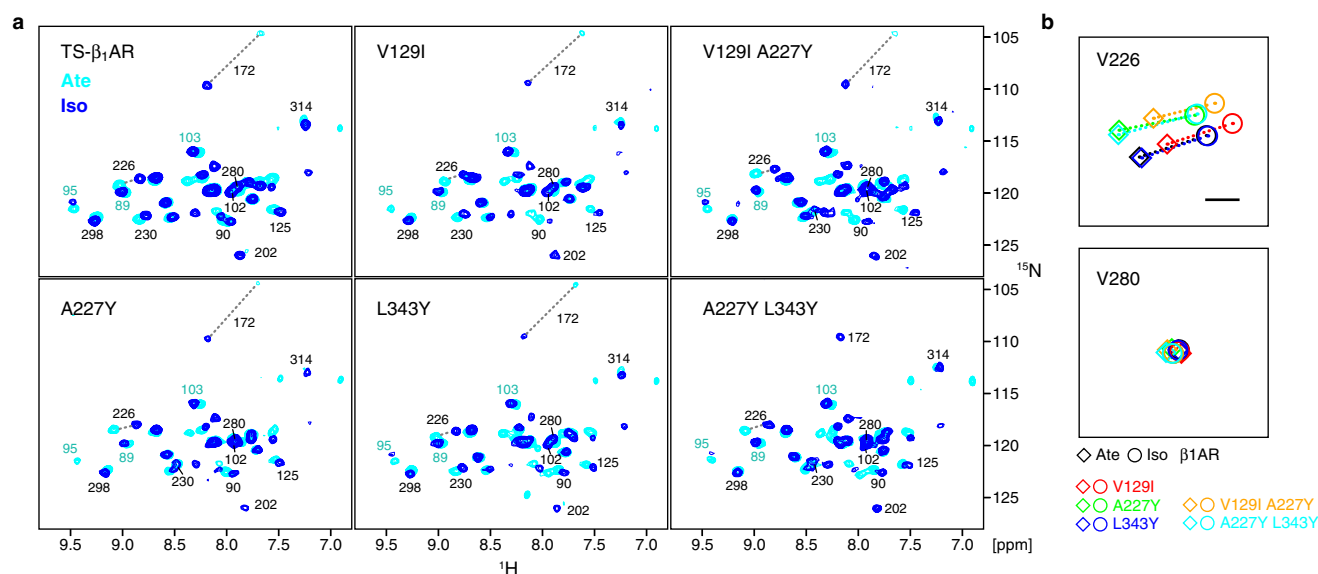
**Extended Data Figure 4 | Evidence for micro- to millisecond dynamics at the ligand entry/exit pathway.** Left, region of the <sup>1</sup>H–<sup>15</sup>N TROSYs showing the V314(6.59) <sup>1</sup>H–<sup>15</sup>N resonance of TS-β<sub>1</sub>AR in the apo and various ligand-bound forms. The resonances of the apo and atenolol-bound forms are severely broadened in the <sup>15</sup>N dimension indicative of backbone dynamics in the micro- to millisecond range. The broadening is not observed for the other more tightly binding ligands. The resonances

for the low affinity ligands, atenolol, isoprenaline and dobutamine, and for the high affinity ligands, alprenolol, carvedilol and cyanopindolol cluster at different positions. Interestingly, the resonance of the apo form clusters with the high affinity ligands, indicative of a similar backbone conformation. Right, <sup>15</sup>N and <sup>1</sup>H line widths of the V314(6.59) resonance for the apo and ligand-bound forms of TS-β<sub>1</sub>AR.



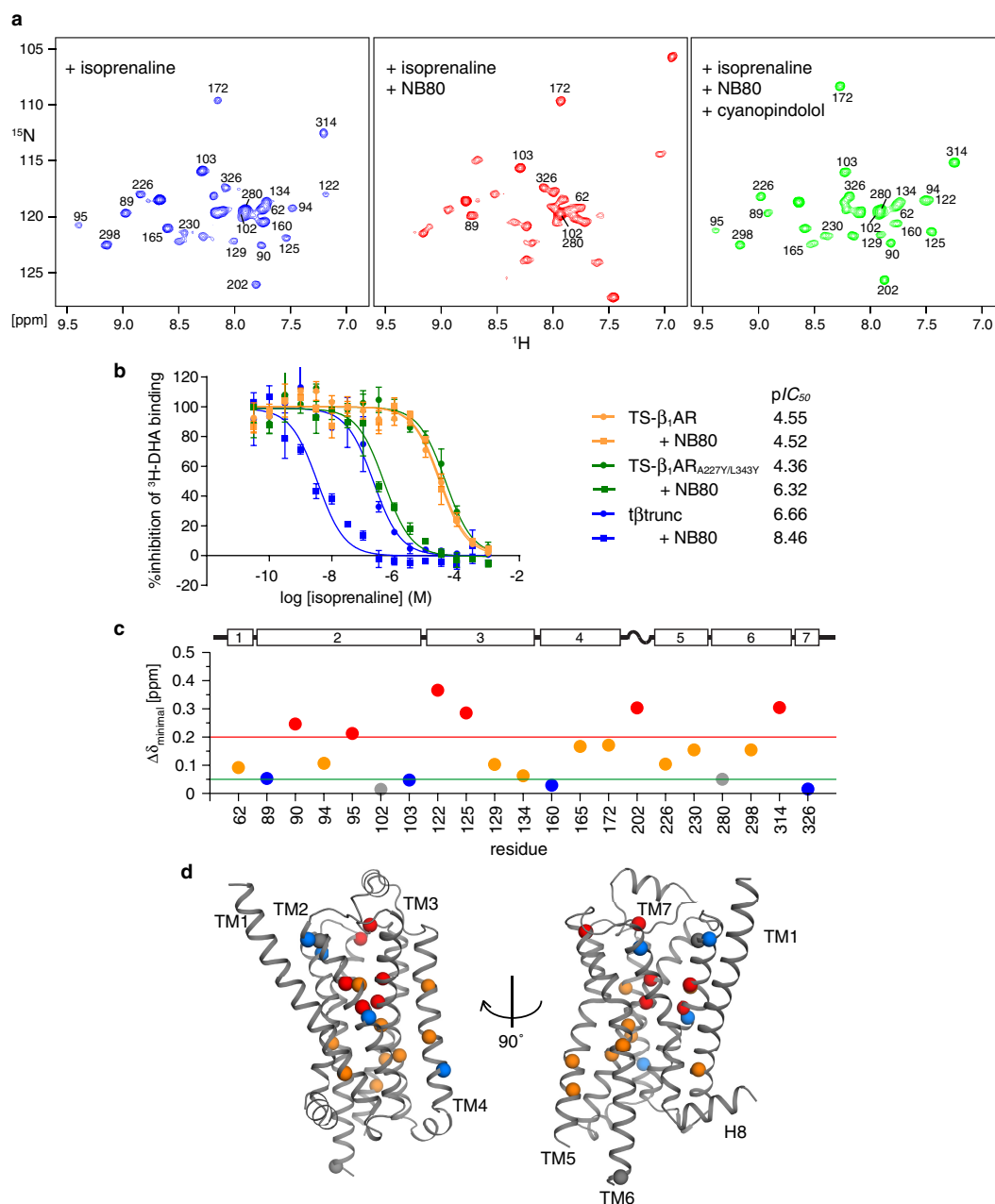
**Extended Data Figure 5 | Isoprenaline-induced G protein activation by  $\beta_1$ AR mutants.** Activation was measured by the change in tryptophan fluorescence upon exchange of GDP for GTP $\gamma$ S in the G $\alpha_i$  subunit of the heterotrimeric G $_i$  protein. The figure shows the time courses of tryptophan fluorescence after the addition of different mutant forms of TS- $\beta_1$ AR in nanodiscs and GTP $\gamma$ S to the heterotrimeric G protein in the presence (red) and absence (black) of the agonist isoprenaline. The increase in fluorescence intensity at time  $t=0$  is caused by the additional fluorescence of the receptor, the nanodisc scaffold protein and the ligand. For the active mutant TS- $\beta_1$ AR(A227Y/L343Y), this initial rise is followed by an

exponential fluorescence increase due to the activation and dissociation of the G protein. The rate of the G $\alpha_i$  activation by TS- $\beta_1$ AR(A227Y/L343Y) (half-life  $\sim 290$  s at 6 nM receptor) is approximately 4 times slower than for the activation by rhodopsin (half-life  $\sim 400$  s at 1 nM receptor) using the same assay. This assumes that the  $K_m$  of G $\alpha_i$  for TS- $\beta_1$ AR(A227Y/L343Y) is of the same scale as for rhodopsin (8.6 nM) and therefore significantly below the (saturating) concentration of G $\alpha_i$  (100 nM). Each mutant was measured three times except for TS- $\beta_1$ AR(A227Y/L343Y) that was measured four times. Representative fluorescence curves from single experiments are shown.



**Extended Data Figure 6 | Observed NMR effects of the reverse mutations of TS- $\beta_1$ AR towards the native  $\beta_1$ AR sequence.** **a**,  $^1\text{H}$ - $^{15}\text{N}$  TROSY spectra of TS- $\beta_1$ AR and several reverse single and double mutants in complex with either atenolol (cyan) or isoprenaline (dark blue). Resonances are marked with assignment information (black, definite; cyan, tentative). **b**, Enlarged regions of the  $^1\text{H}$ - $^{15}\text{N}$  correlation spectra showing only the resonance positions of V226(5.57) and V280(6.25) for all mutants in both ligand-receptor complexes. The black bar represents 0.1 p.p.m. in  $^1\text{H}$  and 1 p.p.m. in  $^{15}\text{N}$ . For all mutants, the resonances for V226(5.57) show efficacy-related chemical shift changes between

atenolol (diamonds) and isoprenaline (circles). For the A227Y mutants TS- $\beta_1$ AR(A227Y) (green), TS- $\beta_1$ AR(V129I/A227Y) (orange) and TS- $\beta_1$ AR(A227Y/L343Y) (cyan), the  $^1\text{H}$ - $^{15}\text{N}$  resonances of V226(5.57) also exhibit an overall shift due to a ring current effect from the introduced Y227(5.58) side chain, which has no structural significance. However, the V129I mutants TS- $\beta_1$ AR(V129I) (red) and TS- $\beta_1$ AR(V129I/A227Y) (orange) exhibit further  $^1\text{H}$ - $^{15}\text{N}$  shifts towards a more active (that is, bent) state of TM5 relative to the mutants that carry the V129(3.40) residue. For all mutants, the V280(6.25) resonances fall basically in identical positions and show no changes between atenolol and isoprenaline.



**Extended Data Figure 7 | Binding of NB80 to agonist-activated TS- $\beta_1$ AR(A227Y/L343Y).** **a**, Sequential addition of isoprenaline (1 mM, blue), NB80 (120  $\mu\text{M}$ , red), and cyanopindolol (1 mM, green) to TS- $\beta_1$ AR(A227Y/L343Y) (120  $\mu\text{M}$ ) followed by  $^1\text{H}$ - $^{15}\text{N}$  TROSY spectra. Addition of an equimolar amount of NB80 changes the spectrum drastically. Except for few resonances (62, 89, 102, 103, 172, 280 and 326), the assignments are not transferable from the isoprenaline complex. The addition of cyanopindolol to this mixture apparently replaces isoprenaline in the receptor and drives the spectrum to the purely cyanopindolol-bound form. These results indicate that TS- $\beta_1$ AR(A227Y/L343Y) can be activated, but that the fully active state requires stabilization with a G protein-like partner at the intracellular site. **b**, Changes in isoprenaline affinity of  $\beta_1$ AR induced by NB80 binding. The isoprenaline affinity was assayed for TS- $\beta_1$ AR (orange), TS- $\beta_1$ AR(A227Y/L343Y) (green), and the truncated native turkey  $\beta_1$ AR receptor (t $\beta$ trunc, blue)<sup>17</sup> by radio-ligand competition using  $^3\text{H}$ -dihydroalprenolol in the presence of saturating amounts of NB80 (10  $\mu\text{M}$ , squares) or its absence (circles). For TS- $\beta_1$ AR(A227Y/L343Y) and t $\beta$ trunc the apparent affinity increases by two log units in the presence of NB80, whereas no change occurs for TS- $\beta_1$ AR. Data points are shown as mean and standard deviation of three independent experiments for each condition. **c**, Combined  $^1\text{H}$  and  $^{15}\text{N}$

chemical shift changes ( $\Delta\delta = (\Delta\delta_{^1\text{H}}^2/2 + \Delta\delta_{^{15}\text{N}}^2/50)^{1/2}$ ) between the isoprenaline-TS- $\beta_1$ AR(A227Y/L343Y) and the NB80-isoprenaline-TS- $\beta_1$ AR(A227Y/L343Y) complexes as a function of residue number. Except for the few assigned peaks in the NB80-isoprenaline-TS- $\beta_1$ AR(A227Y/L343Y) complex, chemical shift deviations are calculated as the minimal difference from the isoprenaline peak position to the nearest peak position in the isoprenaline-NB80 complex. Note that these minimal chemical shift deviations present a lower boundary for the deviations that could be obtained from a full assignment of the isoprenaline-NB80 complex. Peak deviations are colour-coded in red ( $\geq 0.2$  p.p.m.), orange (0.2–0.05 p.p.m.), blue ( $< 0.05$  p.p.m.), and grey (ambiguity from peak overlap). **d**, Representation of the chemical shift response to nanobody binding given in **c** on the structure of  $\beta_1$ AR. Valines in TM1 and TM7 do not show a large chemical shift response. Due to overlap the response for V280(6.25) in TM6 is unclear. Valines in TM2 show a strong response around the sodium binding region near residue D87(2.50), but valines at its extracellular side are not responsive. In general, valine resonances at the extracellular sides of TM3–TM6 are severely shifted. This indicates that the NB80 binding at the intracellular side has a long-range effect on the extracellular side near the ligand binding site.



**Extended Data Table 1 | Pharmacological and geometrical properties of the  $\beta_1$ AR ligands used in this study**

	pharmacology				geometry	
	$pK_D$ $\beta 36\text{-m}23^*$	$pK_i$ $TS\text{-}\beta_1\text{AR}^\dagger$	$pIC_{50}$ $TS\text{-}\beta_1\text{AR}^\ddagger$	Gs efficacy $\dagger\beta\text{trunc}^\S$	insertion depth <sup>  </sup>	tail volume <sup>¶</sup>
cyanopindolol	$10.01 \pm 0.11$	$>8.4 \pm 0.3^\#$	$>8.1 \pm 0.1^\#$	$39 \pm 1$	7.22	90.11
carvedilol	$8.72 \pm 0.09$	$>8.3 \pm 0.3^\#$	$>7.9 \pm 0.1^\#$	$12 \pm 0.4$	7.62	163.47
alprenolol	$7.35 \pm 0.07$	$7.6 \pm 0.3$	$7.2 \pm 0.1$	$31 \pm 2$	7.25	73.87
dobutamine	$5.49 \pm 0.03$	$5.0 \pm 0.3$	$4.6 \pm 0.1$	$112 \pm 7$	8.46	170.34
isoprenaline	$5.16 \pm 0.05$	$4.9 \pm 0.3$	$4.6 \pm 0.2$	100	7.92	73.87
atenolol	$4.29 \pm 0.11$	$5.0 \pm 0.3$	$4.6 \pm 0.2$	$2.1 \pm 1$	ND <sup>°</sup>	73.87

\*Values for binding affinity correspond to the  $\beta 36\text{-m}23$  construct, which contains C- and N-terminal deletions and six thermostabilizing mutations<sup>10</sup>. The  $pK_D = -\log K_D$  [M] was obtained from [<sup>3</sup>H] CGP12177 whole-cell binding assays in stably expressed CHO cells<sup>17</sup>.

<sup>†</sup> $pK_i$  values were obtained from the  $pIC_{50}$  values given in column 3 by adding  $\log(1 + [\text{lig}]/K_D) = 0.366$  as a correction for the concentration of the [<sup>3</sup>H]dihydroalprenolol (<sup>3</sup>H-DHA) ([lig] = 20 nM) and its determined dissociation constant ( $K_D = 15.1 \pm 7$  nM, see Methods). Errors are obtained by error propagation from the errors of the  $pIC_{50}$  values and  $K_D$ .

<sup>‡</sup> $pIC_{50}$  values were determined by the radioligand inhibition assay described in the Methods section using <sup>3</sup>H-DHA. Data are given as mean and standard deviation of three independent experiments. <sup>§</sup>Values for efficacy towards Gs-mediated signalling as determined by adenylyl cyclase activation of the truncated native turkey  $\beta_1$ AR receptor ( $\dagger\beta\text{trunc}$ )<sup>17</sup>. The efficacies are indicated as the percentage of the (maximal) isoprenaline response obtained from <sup>3</sup>H-cAMP accumulation assays in CHO cells.

<sup>||</sup>The insertion depth (Å) of the ligand was taken as the distance between the  $\beta$ -carbon atom of the ligand amino group and the amide nitrogen atom of V125 (V117 for  $\beta_2$ AR) in the crystal structures of turkey  $\beta_1$ AR in complexes with isoprenaline (PDB ID: 2Y03), dobutamine (PDB ID: 2Y00), carvedilol (PDB ID: 4AMJ), and cyanopindolol (PDB ID: 4BVN) as well as of human  $\beta_2$ AR in complex with alprenolol (PDB ID: 3NYA).

<sup>¶</sup>The tail volumes (Å<sup>3</sup>) were calculated by the Molinspiration Property Calculation Service (<http://www.molinspiration.com/cgi-bin/properties>) for the tail group including the amino moiety.

<sup>#</sup>The  $pK_i$  and  $pIC_{50}$  are limited by the  $K_D$  and concentration of the radioactive tracer ligand as well as by the concentration of the receptor in the competition assay.

<sup>°</sup>Not determined, no crystal structure available.

Extended Data Table 2 | Sequence-specific assignment  $^1\text{H}$ – $^{15}\text{N}$  valine resonances

Assigned residue	Position	Point mutation	Spectra with ligands	Further information
V62	1.53	V62A	Ate, Apo, Alp	HN(CO)
V90	2.53	V90A	Ate, Alp, Car, Cya	
V94	2.56	V94A	Alp, Car, Cya	
V102	2.64	V102A	Ate, Apo, Alp, Iso, Dob, Car, Cya	
V122	3.33	V122A	Alp, Car, Cya	Pred
V125	3.36	V125A/I	Alp(A), Car(A), Ate(I), Apo(I)	
V129	3.40	V129I	Ate	
V134	3.45	V134A	Ate	
V160	4.44	V160A	Ate	
V172	4.56	V172A	Ate, Alp	
V202	ECL	V202A	Alp	
V226	5.57	V226A	Alp, Car, Cya	
V230	5.61	V230A	Alp	
V280	6.25	V280A	Alp, Car, Cya	
V298	6.43	V298A	Ate	HN(CO)
V314	6.59	V314A	Alp, Car, Cya	
V326	7.36	V326A	Alp	
V89	2.52	-	Ate, Apo, Alp, Iso, Dob, Car, Cya	HB, NM
V95	2.57	-	Ate, Apo, Alp, Iso, Dob, Car, Cya	HB, NM, HN(CO)
V103	2.65	-	Ate, Apo, Alp, Iso, Dob, Car, Cya	HB, NM, HN(CO)
V165	4.49	-	Ate, Apo, Alp, Iso, Dob, Car, Cya	NM

Ate, atenolol; Alp, alprenolol; Iso, isoprenaline; Dob, dobutamine; Car, carvedilol; Cya, cyanopindolol

HN(CO): information from 2D-filtered HN(CO) of [ $^{15}\text{N}$ ,  $^{13}\text{C}$ ]Gly,Val-, [ $^{15}\text{N}$ ]Leu-labelled receptor

Pred: ShiftX2  $^{15}\text{N}$  chemical shift prediction<sup>31</sup>

HB: expected  $^1\text{H}$  shift based on hydrogen bond length analysis

NM: peak shift due to mutation of neighbouring residue

Assignments for V89(2.52), V95(2.57), V103(2.65), V122(3.33), and V165(4.49) are tentative. V51(1.42), V52(1.43), V56(1.47), V60(1.51), V309(6.54), V312(6.57), and V320(ECL) have not been assigned.

**Extended Data Table 3 | Summary of the NMR response to ligands, G protein activation and NB80 binding data for the original TS- $\beta_1$ AR construct and various reverse mutants towards the native  $\beta_1$ AR sequence**

mutant name	residue			V226 response*	V280 response*	G protein activation <sup>†</sup>	NB80 binding <sup>‡</sup>	$T_m$ [°C] <sup>§</sup>
	129	227	343					
original TS- $\beta_1$ AR	V	A	L	+	-	-	- (4.55/4.52)	58.9 +/- 0.6
TS- $\beta_1$ AR <sub>V129I</sub>	<b>I</b> <sup>  </sup>	A	L	+	-	-	N.D. <sup>¶</sup>	58.6 +/- 0.2
TS- $\beta_1$ AR <sub>A227Y</sub>	V	<b>Y</b>	L	+	-	-	-	54.9 +/- 0.1
TS- $\beta_1$ AR <sub>L343Y</sub>	V	A	<b>Y</b>	+	-	-	-	54.7 +/- 0.3
TS- $\beta_1$ AR <sub>V129I/A227Y</sub>	<b>I</b>	<b>Y</b>	L	+	-	-	N.D.	54.4 +/- 0.1
TS- $\beta_1$ AR <sub>A227Y/L343Y</sub>	V	<b>Y</b>	<b>Y</b>	+	-	+	+ (4.36/6.32)	47.8 +/- 0.2

\*Determined from the shifts of the  $^1\text{H}$ - $^{15}\text{N}$  resonances in response to binding of atenolol and isoprenaline (see Extended Data Fig. 6B).

<sup>†</sup>Determined from the  $G_i$  activation assay described in Extended Data Fig. 5.

<sup>‡</sup>Determined from NMR experiments by mixing of isoprenaline-activated TS- $\beta_1$ AR with NB80. The  $^1\text{H}$ - $^{15}\text{N}$  spectra are shown for TS- $\beta_1$ AR(A227Y/L343Y) in Extended Data Fig. 7. The spectra of the other TS- $\beta_1$ AR mutants showed no response to addition of NB80 up to equimolar concentrations of  $\sim 100\ \mu\text{M}$  for both proteins. Values in parentheses show  $\text{pIC}_{50}$  for isoprenaline binding in the absence or presence of NB80, respectively (see Extended Data Fig. 7).

<sup>§</sup>Melting temperature  $T_m$  of the apo form receptor determined by the CPM thermal shift assay.

<sup>||</sup>Native residues are highlighted in bold.

<sup>¶</sup>Not determined.

## CORRIGENDUM

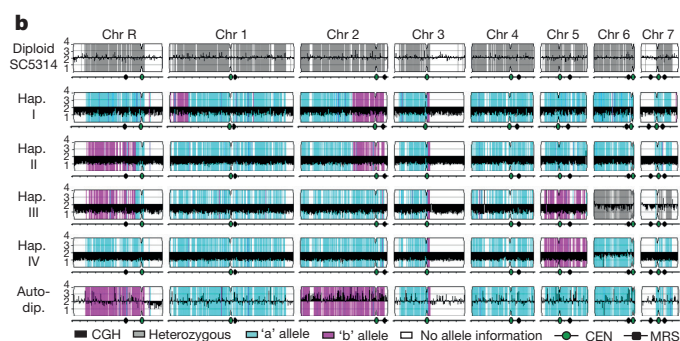
doi:10.1038/nature16134

### Corrigendum: The ‘obligate diploid’ *Candida albicans* forms mating-competent haploids

Meleah A. Hickman, Guisheng Zeng, Anja Forche, Matthew P. Hiraikawa, Darren Abbey, Benjamin D. Harrison, Yan-Ming Wang, Ching-hua Su, Richard J. Bennett, Yue Wang & Judith Berman

*Nature* **494**, 55–59 (2013); doi: 10.1038/nature11865

In this Article, there were two errors in allele assignments. In the legend and colour key to Fig. 1b, the allele identity (‘a’ and ‘b’) is inadvertently reversed. Figure 1 of this Corrigendum shows the corrected panel b and legend for the original Fig. 1. A similar error is present in the colour key to Supplementary Fig. 5 (its legend is correct). The Supplementary Information to this Corrigendum shows the corrected Supplementary Fig. 5.



**Figure 1 | This figure shows the corrected panel b of the original Fig. 1.** SNP/CGH array analysis of indicated strains (left) showing copy number ( $\log_2$  ratio, black) and SNP allele information (grey, heterozygous; cyan, allele ‘a’; magenta, allele ‘b’; white, no SNP data).

**Supplementary Information** is available in the online version of the Corrigendum.



## CORRIGENDUM

doi:10.1038/nature16179

### **Corrigendum: Human body epigenome maps reveal noncanonical DNA methylation variation**

Matthew D. Schultz, Yupeng He, John W. Whitaker, Manoj Hariharan, Eran A. Mukamel, Danny Leung, Nisha Rajagopal, Joseph R. Nery, Mark A. Urich, Huaming Chen, Shin Lin, Yiing Lin, Inkyung Jung, Anthony D. Schmitt, Siddarth Selvaraj, Bing Ren, Terrence J. Sejnowski, Wei Wang, & Joseph R. Ecker

*Nature* **523**, 212–216 (2015); doi:10.1038/nature14465

The y-axis label in Fig. 3f of this Letter should read '[mCAC/CAC]/[mCAG/CAG]' rather than '[mCAC/CAC]/[mCAC/CAG]'. In addition, in the description of Fig. 4c in the main text, 'tissue' and 'individual' were swapped; it should read, "Of the ASM events that varied, 4.1–7.5% and 54.5–70.0% were tissue- and individual-variable, respectively." These errors have been corrected in the online versions of the paper.

# CORRECTIONS & AMENDMENTS

---

## ERRATUM

doi:10.1038/nature16182

### **Erratum: Differential responses to lithium in hyperexcitable neurons from patients with bipolar disorder**

Jerome Mertens, Qiu-Wen Wang, Yongsung Kim, Diana X. Yu, Son Pham, Bo Yang, Yi Zheng, Kenneth E. Diffenderfer, Jian Zhang, Sheila Soltani, Tameji Eames, Simon T. Schafer, Leah Boyer, Maria C. Marchetto, John I. Nurnberger, Joseph R. Calabrese, Ketil J. Oedegaard, Michael J. McCarthy, Peter P. Zandi, Martin Alda, Caroline M. Nievergelt, The Pharmacogenomics of Bipolar Disorder Study, Shuangli Mi, Kristen J. Brennand, John R. Kelsoe, Fred H. Gage & Jun Yao

*Nature* **527**, 95–99 (2015); doi:10.1038/nature15526

In this Letter, author ‘Martin Alda’ was incorrectly listed as ‘Martin Alba’ owing to a production error. In addition, ‘Ketil J. Ødegaard’ should have been listed as ‘Ketil J. Oedegaard’ and their affiliation should have been ‘Division of Psychiatry, Haukeland University Hospital and Department of Clinical Medicine, Section for Psychiatry, University of Bergen and K.G. Jebsen Centre for Research on Neuropsychiatric Disorders, 5021 Bergen, Norway’ rather than ‘Department of Psychiatry, University of Bergen, Bergen 5020, Norway’. These have been corrected in the online versions of the paper.

# CAREERS

**DIVERSITY ADVOCATE** Astrophysicist sends message to US Supreme Court **p.245**

**FACULTY MEMBERS** Tips for effective mentoring [go.nature.com/6g4bnp](http://go.nature.com/6g4bnp)

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)

DMITRI GUZHANIN/GETTY



## COLUMN

### A bridge to business

PhD holders should not underestimate their value to industry and the business sector, says **Peter Fiske**.

The glossy poster at my university career planning and placement centre both intrigued and perplexed me: “PhDs: come learn about a career in management consulting — recruiting reception tonight!”

As a PhD student in geology, I was dimly aware of the name of the consulting firm that had organized the event; one of my friends had accepted a position with the firm after he graduated. I had heard that he was earning a great salary and travelling a lot, and that he was aiming to go back for a master's degree in business administration (MBA). Until I saw the poster, I had no idea that a strategic-management-consulting firm would even consider hiring non-MBA types, let alone recruit PhDs specifically.

I applied for a position and landed a series of interviews that culminated in a day of them at the company's offices in Los Angeles, California. I did not receive an offer in the end, but it was an illuminating experience. Mostly, I was surprised that someone with a freshly minted PhD could immediately earn US\$160,000. And that was in 1994.

Fast forward 22 years. Now, as the chief executive of a US technology company, I have recruited PhDs for both technical and business positions. I have found across all scientific disciplines that those with doctoral degrees possess many of the skills that are in highest demand in today's economy. If you have earned a PhD, you know, for example, how to analyse data. You also understand how to examine those results to gain insights. In some important ways, you are better prepared than MBA holders to make valuable contributions to the business world. You have learned resilience in the face of uncertainty and with limited resources.

Yet you and many other PhD graduates — along with the programmes that trained you — remain largely unaware of or uninterested in opportunities outside academia. In turn, only a few companies, such as the consulting firm whose recruiting poster I saw 20 years ago, have pre-emptively recognized the value that you and your colleagues can bring — and they are reaping a harvest of talent as a result.

You are doing yourself a disservice. As a doctoral degree-holder, you need to appreciate your degree programme for the transferable skills that it confers (see “Top transferable skills for business”), and recognize that those skills provide you with significant and immediate advantages over your business-school counterparts. You do not need a business degree or substantial extra training to secure satisfying and highly ►

► compensated work in the business sector.

Need more convincing? Let's examine the differences between the training components of each degree programme.

In a typical MBA scheme, students listen to lectures, work on projects and learn about business-friendly topics such as economics, finance, accounting, organizational behaviour and law. Fundamentally, the MBA is a conventional, curriculum-led educational experience. Although internships and student-led organizations provide opportunities for direct work experience, success in business school is largely a function of earning good grades. Students graduate after two years or so, and top performers can land jobs that start at \$150,000 or more.

Most doctoral programmes also require a large body of course work. But as a PhD student, you probably spent more time 'doing', rather than studying or passively listening. You conducted original scientific research, which may have included fieldwork, you taught younger students and you were possibly involved with at least one manuscript.

#### DATA DELUGE

Science PhD students particularly benefit from near-constant immersion in emerging technologies and, especially, in data analysis and hypothesis testing. This is a huge advantage in the business world: every industrial sector and

type of business is more reliant on technology and data now than ever before. And in today's economy, the volume and complexity of the information that businesses use require managers and executives to understand the basics of data analysis and statistics — and to have a deeper ability to construct and test hypotheses and to build and validate models.

And science-PhD holders have a further advantage. Not only is technological innovation a key to the development of products and services, but the tools that are used to run businesses have science and technology embedded in them. Customer-relations management software, for example, uses advanced statistical algorithms that analyse trends and identify important variables such as user behaviour and pricing trends.

MBA graduates do not routinely encounter such training or confront these requirements in their degree programme, whereas they are an integral component of the PhD programme.

In short, the PhD experience is much more like the real world of business than is the MBA programme. If educational institutions were to train MBA students in the same way that they train PhD students, they would require business students to complete all the required course work — and to launch a successful business.

Given that the PhD is such a powerful

training foundation for success in the private sector, one might presume that this value would be reflected in the salaries of PhD graduates as compared to those of their MBA counterparts. Unfortunately, the opposite is true.

For example, a PhD-trained biochemist with five years' work experience after their PhD who is employed as a 'PhD biochemist III' (a typical job title in industry) in the United States has an average annual salary of roughly \$68,000, according to salary.com, a major salary database. In contrast, a person with the title 'market-research supervisor III', who also has five years' work experience and an MBA, is likely to earn \$91,000.

Both jobs require an ability to analyse data, synthesize technical information, perform tests to validate hypotheses and supervise a team. But whereas the market-research supervisor almost certainly could not walk into the PhD biochemist's position, the biochemist could easily perform all the duties of the research supervisor — with just a small amount of tailored training.

#### BRIDGE THE GAP

But therein lies the glitch. If a PhD provides graduates with such a range of strengths and transferable skills, why don't doctorate holders immediately command higher salaries? Part of the problem is that PhD students do not routinely receive such tailored training in the skills that are highly valued in the business sector. Practical professional skills such as negotiation, communication, business strategy, basic economics and marketing are all abundantly represented in the curricula of leading MBA programmes.

Yet only a handful of research universities train PhD students in the basics of business. Some institutions, such as Princeton University in New Jersey, have begun to provide professional-development curricula to those in PhD programmes. It is likely that graduates from these places will face better professional and economic prospects than those without such training, because they are better prepared to compete for a broad range of professional opportunities.

There may be a disincentive for science and engineering education programmes (and the agencies that fund them) to support the implementation of such professional-development classes or initiatives. The research enterprise thrives on an abundant reservoir of highly talented, highly motivated PhD students and recent graduates, and does not want to lose them to the business sector.

Greater competition for PhD-trained talent from the private sector, however, would invariably raise salaries for entry-level PhD graduates. Furthermore, science-funding agencies rely on PhD programmes and individual advisers to provide the bulk of professional development to their students. Many research faculty members lack sufficient

### THE TALENT WITHIN

#### *Top transferable skills for business*

If you have earned a science PhD, you were probably told by mentors, advisers and career-development specialists that you will need to develop a lot of new skills to succeed in any sector outside academia. But your PhD programme has already conferred many skills that are important, even crucial, in the business world, and that are comparable to — and in some cases superior to — the talents acquired in a graduate-level business programme. Here are some examples.

● **Data analysis** You were trained to gather, evaluate, synthesize and present data, and to uncover relationships, correlations and trends. The business world increasingly relies on the same methodologies to develop strategies and identify opportunities.

● **Resourcefulness** You probably had to create experiments, methodologies and analyses with limited resources and under tight time constraints. Successful business people are often challenged to develop a product or service while facing the same difficulties.

● **Technological awareness** You were trained to understand the fundamentals of a range of technologies. Many of these

technologies are at the heart of products and services in the private sector.

● **Resilience** You may have encountered unexpected setbacks in your research or studies, yet powered through to reach your goals. This resilience in the face of challenge often separates the most successful entrepreneurs from the rest.

● **Project management** Completing a PhD typically requires the coordination and scheduling of disparate resources and individuals — as well as thinking through all aspects of a complex project or activity. The same course of action is a core component of the business world.

● **Problem-solving** You had to use novel thinking and innovative frames of reference to identify and solve technical problems. The ability to reframe problems to identify novel solutions is a key skill in business.

● **English proficiency** You are probably skilled in English, the most prevalent language of international business.

● **Written communication** PhD holders often have extensive experience in writing and describing complex ideas and methodologies. Effective written communication is crucial to business success. **PF**



industry or professional-development experience to provide the sort of training that MBA students receive as a customary part of their degree.

If a greater proportion of PhD graduates were to transition into the business sector, however, it could create pressure on universities to provide such training and on funding agencies to require it.

In the meantime, this training gap can easily be filled by taking short courses or programmes in business skills. Many are available online as massive open online courses, and some offer certifications.

Still, even more so than the paucity of professional-development programmes for PhD students, the greatest barrier to a high-paying position in the business sector is your personal beliefs about what you are qualified to pursue.

As someone with a PhD background, you probably view yourself foremost — even solely — as

a research scientist. **“The business world is full of delicious, complex, intellectually exciting problems.”**

The broad set of valuable transferable skills that you developed while in graduate school go largely unrecognized and unarticulated within the academy. Most PhD graduates restrict their job searches to what they feel qualified to do, rather than exploring what they are capable of doing.

And as a PhD graduate, you may believe that academia is the only sector in which you can enjoy intellectual freedom and work on challenging problems. This is incorrect. The business world is full of delicious, complex, intellectually exciting problems. Their resolution can yield enormous value, both to those who solve the problems and to society.

Some people are beginning to recognize this reality. Graduate students and postdocs are organizing their own professional-development programmes, sometimes recruiting business-school professors to help them (see *Nature* **485**, 269–270; 2012).

And more businesses, especially technology-related companies, are either being launched by graduates of doctoral programmes or have a PhD holder on the founding team.

As more PhD students and graduates learn about these opportunities from their brethren, we can expect further interest in, and greater pursuit of, business careers by doctorate holders. MBAs, beware. ■

**Peter Fiske** is chief executive of PAX Water Technologies in Richmond, California, and author of *Put Your Science to Work* (American Geophysical Union, 2001).

## TURNING POINT

# Diversity ruling

*Lia Corrales, a postdoctoral researcher at the Massachusetts Institute of Technology (MIT) in Cambridge, is fighting for equality in astronomy — a field that has been plagued by allegations of harassment and low diversity. Last December, Corrales and her colleagues wrote an open letter to the US Supreme Court after some judges questioned whether a federal policy known as ‘affirmative action’ has helped people from minority groups to become scientists.*

### What happened at the Supreme Court?

The justices were debating the merits of affirmative action, which aims to protect against discrimination on the basis of race or gender, in higher education. Justice Antonin Scalia noted that some people contend that the policy does not benefit African Americans because they do better at a less-advanced, slower-track school. And the chief justice, John G. Roberts Jr, asked what “unique perspective” a minority student could bring to a physics class.

### Why did you write an open response?

I was angry that the justices’ comments implied that affirmative action has no value, and knew that friends were upset, too. Astrophysicist Josh Tan, a colleague of mine, and I discussed writing a letter because, as physicists and astronomers who care about diversity, we’re not doing our job if we don’t stand up for affirmative action. In the end, five of us wrote it. I put the letter online at [eblur.github.io/scotus](http://eblur.github.io/scotus) and mailed printed copies to the justices on the case. We gathered more than 2,400 signatures.

### Describe your pursuit of astronomy.

As a kid, I loved black holes and thought Stephen Hawking was the coolest, so it seemed natural to study physics and astronomy. While I was an undergraduate at Harvey Mudd College in Claremont, California, roughly one-third of the students there — but probably half of my physics class — were women. And there were plenty of women at Columbia University in New York, where I earned my PhD studying interstellar dust. But when I came to MIT, where only 3 or 4 of about 30 postdocs at my institute were women, it became apparent how few of ‘me’ there were.

### When did you first work on diversity issues?

As an undergraduate. The diversity office at Harvey Mudd College asked me to help start a branch of the US Society of Hispanic Professional Engineers. But I also wanted to make a difference to aspiring university students, so we connected with a preparatory programme



for students who would be the first in their family to attend university. I now realize that I’m passionate about this issue because of my own Latin-American background and my angst about affirmative action.

### You are part of a Facebook group called Equity and Inclusion in Physics and Astronomy. Can you tell us more about it?

I’m not a founder, but it started out as a ‘Women in Astronomy’ group on Facebook, and evolved into something that was about more than just gender. There are many ways in which people experience discrimination, so now it’s more of an umbrella group to increase diversity.

### What else have you been involved with?

Last summer, an Inclusive Astronomy conference was held at Vanderbilt University in Nashville, Tennessee — an institution where several faculty members have tackled diversity issues. It was an amazing event. There was so much acceptance that we must confront stereotypes and move beyond the ways in which we allow or contribute to discrimination.

### Are things beginning to change in astronomy?

There is a critical mass of people who are talking about diversity and inclusivity. One recommendation from the conference was to stop the requirement for minimum test scores for admission to US graduate schools. This stance has already been adopted by the American Astronomical Society — a step forward. There’s still a long way to go, but I feel more inspired and empowered than I have in a really long time. ■

### INTERVIEW BY VIRGINIA GEWIN

This interview has been edited for length and clarity.

industry or professional-development experience to provide the sort of training that MBA students receive as a customary part of their degree.

If a greater proportion of PhD graduates were to transition into the business sector, however, it could create pressure on universities to provide such training and on funding agencies to require it.

In the meantime, this training gap can easily be filled by taking short courses or programmes in business skills. Many are available online as massive open online courses, and some offer certifications.

Still, even more so than the paucity of professional-development programmes for PhD students, the greatest barrier to a high-paying position in the business sector is your personal beliefs about what you are qualified to pursue.

As someone with a PhD background, you probably view yourself foremost — even solely — as

a research scientist. **“The business world is full of delicious, complex, intellectually exciting problems.”**

The broad set of valuable transferable skills that you developed while in graduate school go largely unrecognized and unarticulated within the academy. Most PhD graduates restrict their job searches to what they feel qualified to do, rather than exploring what they are capable of doing.

And as a PhD graduate, you may believe that academia is the only sector in which you can enjoy intellectual freedom and work on challenging problems. This is incorrect. The business world is full of delicious, complex, intellectually exciting problems. Their resolution can yield enormous value, both to those who solve the problems and to society.

Some people are beginning to recognize this reality. Graduate students and postdocs are organizing their own professional-development programmes, sometimes recruiting business-school professors to help them (see *Nature* **485**, 269–270; 2012).

And more businesses, especially technology-related companies, are either being launched by graduates of doctoral programmes or have a PhD holder on the founding team.

As more PhD students and graduates learn about these opportunities from their brethren, we can expect further interest in, and greater pursuit of, business careers by doctorate holders. MBAs, beware. ■

**Peter Fiske** is chief executive of PAX Water Technologies in Richmond, California, and author of *Put Your Science to Work* (American Geophysical Union, 2001).

## TURNING POINT

# Diversity ruling

*Lia Corrales, a postdoctoral researcher at the Massachusetts Institute of Technology (MIT) in Cambridge, is fighting for equality in astronomy — a field that has been plagued by allegations of harassment and low diversity. Last December, Corrales and her colleagues wrote an open letter to the US Supreme Court after some judges questioned whether a federal policy known as ‘affirmative action’ has helped people from minority groups to become scientists.*

### What happened at the Supreme Court?

The justices were debating the merits of affirmative action, which aims to protect against discrimination on the basis of race or gender, in higher education. Justice Antonin Scalia noted that some people contend that the policy does not benefit African Americans because they do better at a less-advanced, slower-track school. And the chief justice, John G. Roberts Jr, asked what “unique perspective” a minority student could bring to a physics class.

### Why did you write an open response?

I was angry that the justices’ comments implied that affirmative action has no value, and knew that friends were upset, too. Astrophysicist Josh Tan, a colleague of mine, and I discussed writing a letter because, as physicists and astronomers who care about diversity, we’re not doing our job if we don’t stand up for affirmative action. In the end, five of us wrote it. I put the letter online at [eblur.github.io/scotus](http://eblur.github.io/scotus) and mailed printed copies to the justices on the case. We gathered more than 2,400 signatures.

### Describe your pursuit of astronomy.

As a kid, I loved black holes and thought Stephen Hawking was the coolest, so it seemed natural to study physics and astronomy. While I was an undergraduate at Harvey Mudd College in Claremont, California, roughly one-third of the students there — but probably half of my physics class — were women. And there were plenty of women at Columbia University in New York, where I earned my PhD studying interstellar dust. But when I came to MIT, where only 3 or 4 of about 30 postdocs at my institute were women, it became apparent how few of ‘me’ there were.

### When did you first work on diversity issues?

As an undergraduate. The diversity office at Harvey Mudd College asked me to help start a branch of the US Society of Hispanic Professional Engineers. But I also wanted to make a difference to aspiring university students, so we connected with a preparatory programme



for students who would be the first in their family to attend university. I now realize that I’m passionate about this issue because of my own Latin-American background and my angst about affirmative action.

### You are part of a Facebook group called Equity and Inclusion in Physics and Astronomy. Can you tell us more about it?

I’m not a founder, but it started out as a ‘Women in Astronomy’ group on Facebook, and evolved into something that was about more than just gender. There are many ways in which people experience discrimination, so now it’s more of an umbrella group to increase diversity.

### What else have you been involved with?

Last summer, an Inclusive Astronomy conference was held at Vanderbilt University in Nashville, Tennessee — an institution where several faculty members have tackled diversity issues. It was an amazing event. There was so much acceptance that we must confront stereotypes and move beyond the ways in which we allow or contribute to discrimination.

### Are things beginning to change in astronomy?

There is a critical mass of people who are talking about diversity and inclusivity. One recommendation from the conference was to stop the requirement for minimum test scores for admission to US graduate schools. This stance has already been adopted by the American Astronomical Society — a step forward. There’s still a long way to go, but I feel more inspired and empowered than I have in a really long time. ■

### INTERVIEW BY VIRGINIA GEWIN

This interview has been edited for length and clarity.



# THE MAN WITH THE SPIDER

*Escape clause.*

BY MOHAMAD ATIF SLIM

“I want out.”  
“I’m sorry, sir?” Surely she had misheard.

“I want to get out.”

The receptionist seemed slightly alarmed. Her eyes stole to the intravenous device on the man’s forearm.

“Sir, are you ... *sober*?” She hesitated. “We do offer a wide range of reversal agents for all classes of pharma, guaranteed not to compromise any subsequent pleasure —”

“But I *am* sober!” Carusi was dismayed to hear the creeping hysteria in his voice. “Please. I’d like to leave.”

“I’m ... not sure if that’s possible, sir.”

“I signed the contract, I know. But I’ve changed my mind. Do whatever you want with me, just not this. I can’t go through with this. I need to see my wife.”

Her eyes again fell on Carusi’s arm.

“Please. If you’re not able to help, can you get someone who can?”

“I ... Certainly, sir. Please have a seat.”

As she retreated into a back room, Carusi gazed anxiously out of the window. The sky was azure over plains of green, and a distant bird soared above snow-flecked mountains. The hyper-realistic projections were designed to deceive; MARTYR was at least 100 kilometres underground.

Like many, Carusi had lapped up the endless supply of stimulants at the beginning. Without sleep, you start to lose count of the days. Time becomes an abstraction of pleasure after pleasure after pleasure. For most, the hour of martyrdom approached with stealth: a brief, errant breeze, then quiet.

But yesterday, Carusi decided to forgo the diafinil. Something gnawed at the back of his mind, a place no pharma could touch. For the first time in perhaps a week, he slept, only to wake up hours later, drenched in sweat. Colomen’s face swam across his vision, and he stifled a cry.

That was when panic, that many-legged spider, took residence in the pit of his stomach.

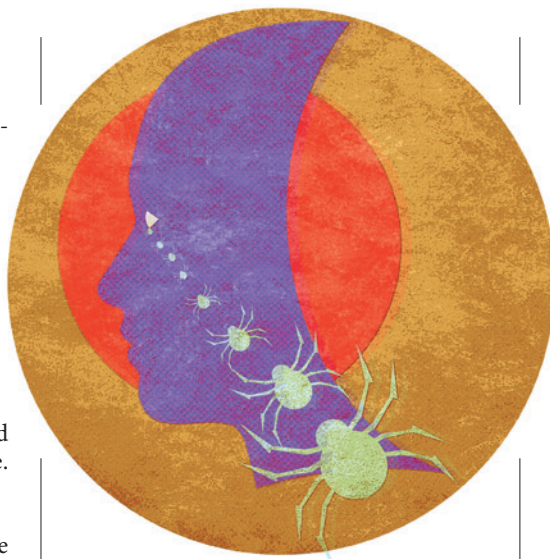
He needed to go home.

How many days had passed since he enlisted?

How many did he have left?

What a fool he was!

He had wanted to lower her mother’s medical cap — palliative, then assisted decline, after 60. Sell the excess. It was the only way they could afford a reproductive



licence before June. They had already missed the sterilization lottery two years in a row. Three would be too lucky.

And Ma didn’t need to know.

But Colomen was stubborn. “We’ll work, Carusi. We’ll *earn* our licence.”

The arguments were long and many, but the last was the worst. Carusi had stormed out. How could Colomen insist on choosing her mother over *him*? Over *their* future?

He might have left for hours. He might have got drunk and seen the ad for MARTYR in the subway: “3D-printing for food will saturate in 2 years — for the sake of the children, enlist as a Martyr!”

And he might have, in the heat of the moment, thought it was a good idea.

Martyrdom! For the sake of the child he would never have.

The registrar didn’t ask questions, even though Carusi was slurring his words. Martyrdom was a right; in exchange, state-sanctioned pleasure for 13 days — boundless, limitless pleasure. And they distort the perception of time too, he’d heard. Make 13 days feel like for ever.

But he’d also heard other things. Rumour orbits MARTYR like flies round a carcass.

The government calls it the solution to over-population: the old for the new, the infirm for the healthy. Cattle are long dead, and oil wells dry; scientists say once the population reaches critical mass, anarchy will prevail — the *children* will suffer.

Yet, others call it foul. *Class-based eugenics*, they say. *Institutionalized nihilism*. They say only the proletariat enlist because society offers them nothing else. That

MARTYR cannot save a corrupt world. That in secret places around the Metropolis, the bourgeoisie scramble to bid at auctions for licences the Martyrs forfeit.

But Carusi didn’t care about politics, or much about the world. He had enlisted because he was angry, petty and spiteful.

He knew that now.

“Mr Onnodig?”

Carusi turned with a start. A well-dressed man beckoned him into the back room, an expression of grave concern on his face.

“I am very sorry, sir,” he began, once they were seated. “I empathize with your predicament, but MARTYR cannot allow for any process reversal. The cogs turn one way.”

Inside, the spider gave a malevolent hiss. “But I was unwell!” Carusi pleaded, then added, shamefully: “And inebriated. I was angry with my wife.”

“I would like to help, but we have no legal provisions.” He handed Carusi a holotablet. “Your contract, sir. Page 254. Clause 537. *An Acute Stress Reaction is not tenable grounds for denying a Citizen his right to Martyrdom.*”

“But what about my rights now?”

“Martyrs have basic rights, but their right to Martyrdom supersedes all else.” He paused. “You must understand — MARTYR only works if everyone plays their part. The government compensates the citizen with a commensurate entitlement to pleasure. In turn, you surrender your licence to live.”

Carusi felt the spider climb into the hollow of his chest, lodging itself beneath his ribs. Its jaws chattered angrily.

He stared at the opposite wall, unable to find comfort in the gaunt figure reflected in its polished wooden panels.

The girl from earlier approached him with a syringe, her gestures gauche.

“Lotazolam, sir. It can help with anxiety.”

Carusi ignored her. “Can I call my wife?”

The man shifted uncomfortably in his seat. “This is very... *unconventional* for us. No one’s ever asked to leave.” The man seemed to debate internally, glancing at the girl. “But we will spare you five minutes, no more.”

They left the man with the spider alone in the room. A minute later, his wife’s face flickered onto the holotablet, and Carusi broke down into sobs. ■

**Mohamad Atif Slim** is a nomadic Malaysian presently based in Hamilton, New Zealand, the City of the Future, where he works as a secretarial diagnostician.

ILLUSTRATION BY JACEY